$< \mathcal{SOG}_k >$: One LLM Token For Explicit Graph Structural Understanding

Jingyao Wu[1], Bin Lu[1]*, Zijun Di[1], Xiaoying Gan[1], Meng Jin[1], Luoyi Fu[1], Xinbing Wang[1], Chenghu Zhou[2]

[1]Shanghai Jiao Tong University (SJTU)
[2]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (IGSNRR)

Introduction: LLMs for Graph Understanding

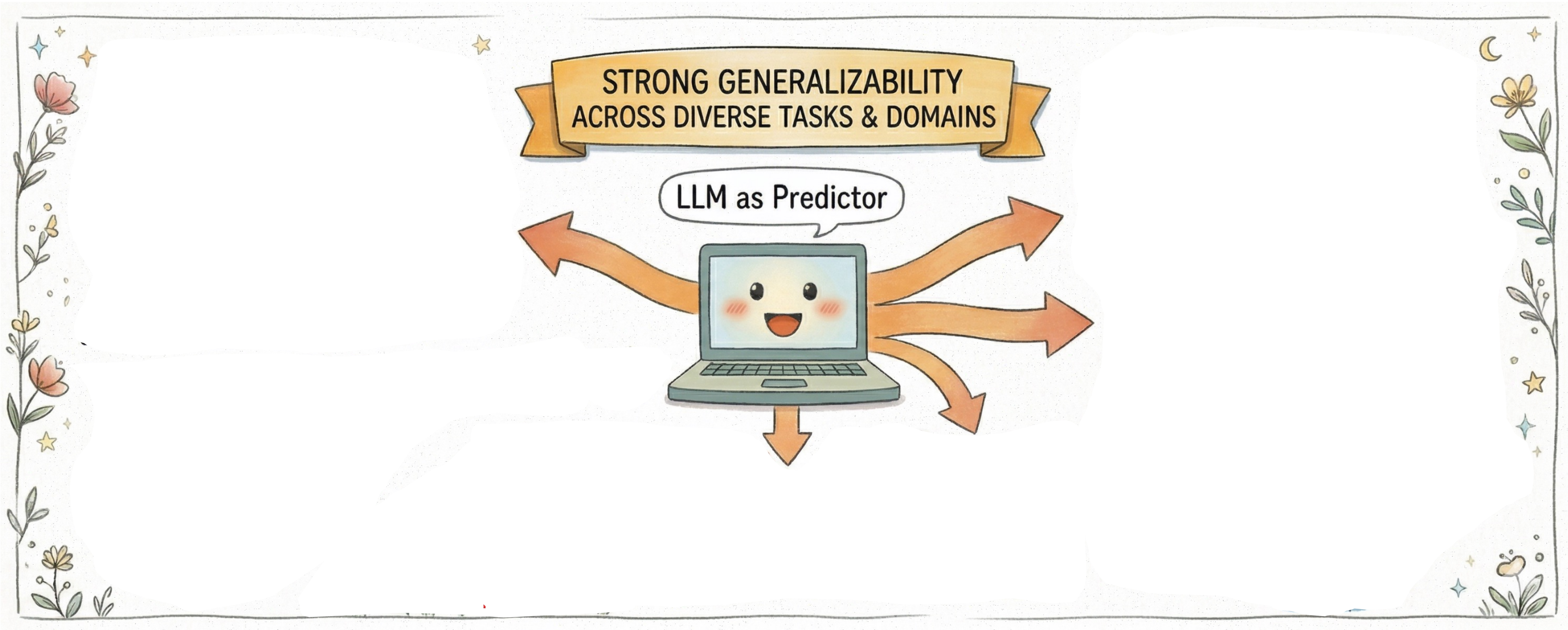
■ The Dominant Paradigm: LLM as Predictor.

- LLM as Predictor gained widespread attention due to its strong generalizability across different level tasks and diverse datasets/domains.

Introduction: LLMs for Graph Understanding

■ **The Dominant Paradigm: LLM as Predictor.**

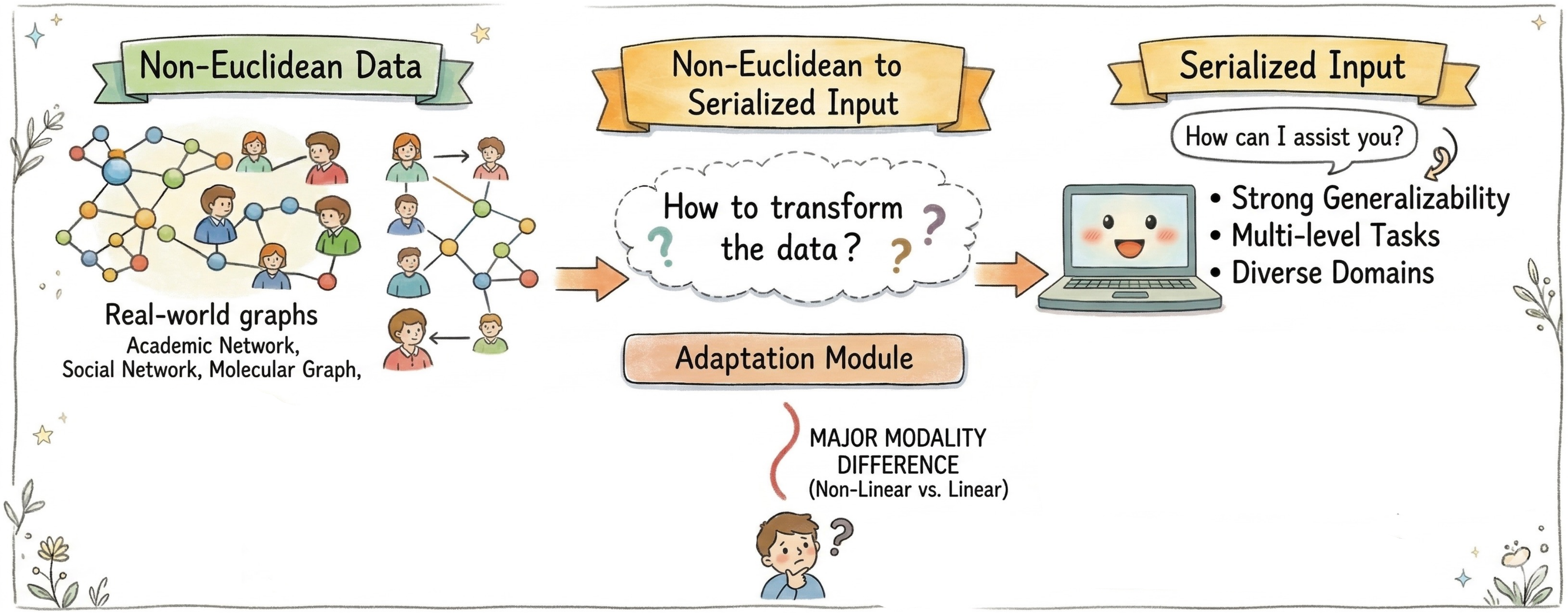
- ➢ LLM as Predictor gained widespread attention due to its strong generalizability across different level tasks and diverse datasets/domains.



Introduction: LLMs for Graph Understanding

■ The Dominant Paradigm: LLM as Predictor.

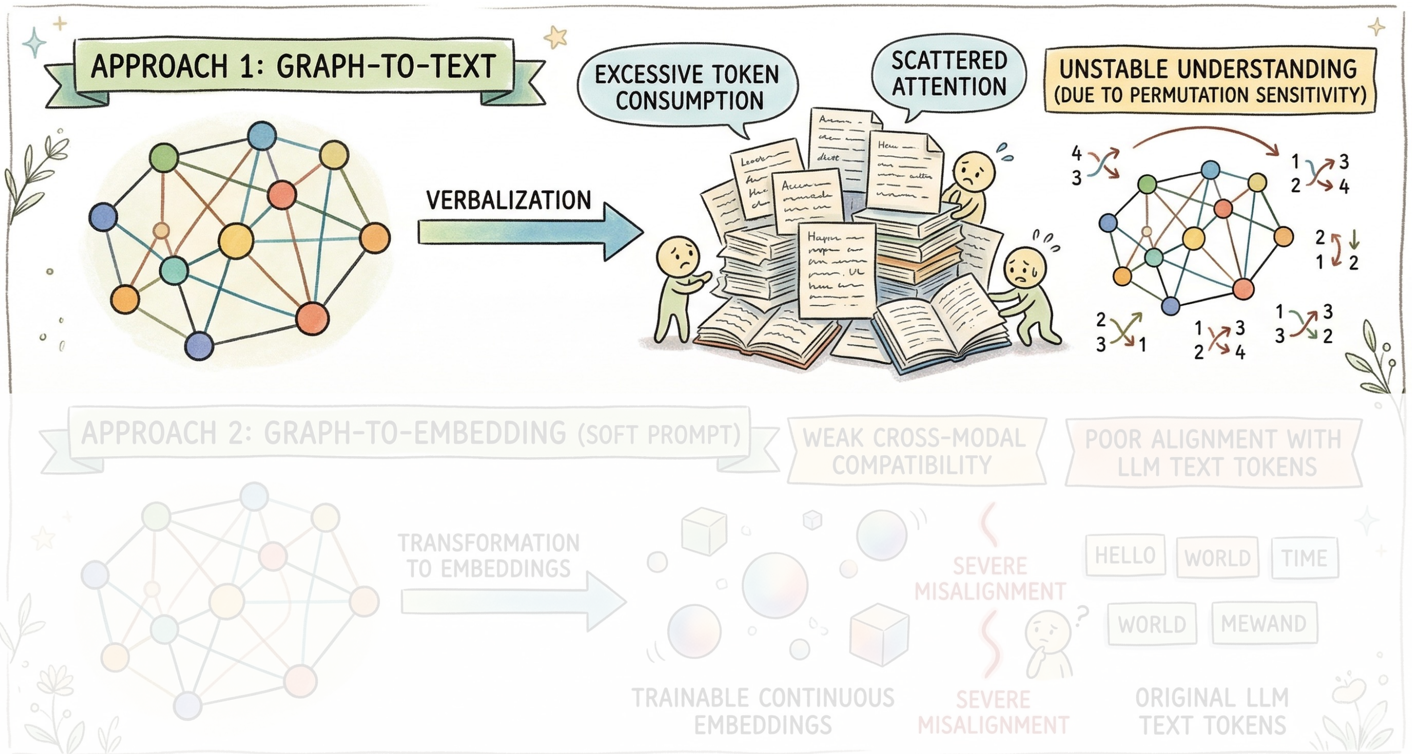
- LLM as Predictor gained widespread attention due to its strong generalizability across different level tasks and diverse datasets/domains.
- The core bottleneck: how to feed **non-Euclidean** graph structure for **serialized** LLM inputs.



Introduction: LLMs for Graph Understanding

■ Limitations of Existing Approaches.

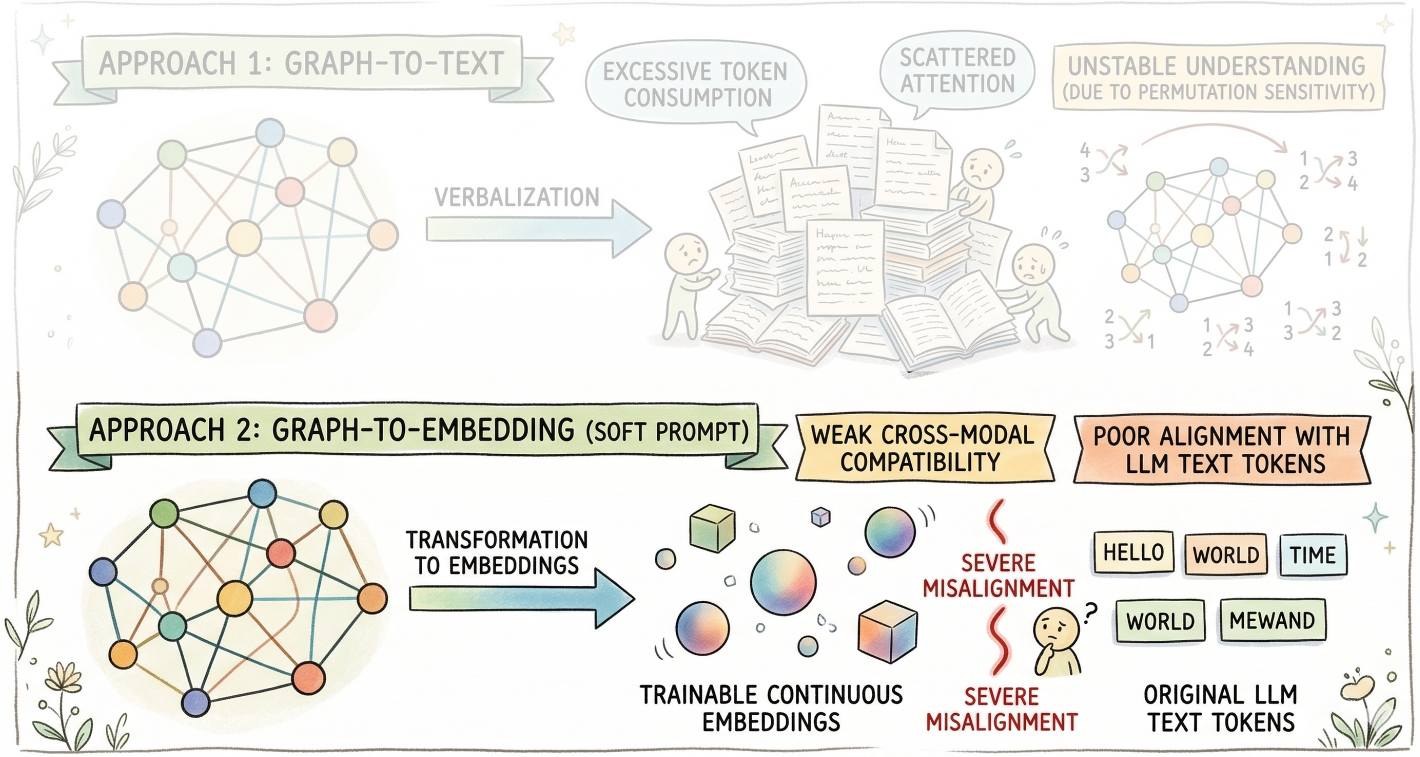
- ➢ **Graph-to-Text**: excessive token consumption, scattered attention and permutation sensitivity.(not concise)
- ➢ **Graph-to-Embedding**: severe misalignment, weak cross-modal compatibility. (not accurate)



Introduction: LLMs for Graph Understanding

■ Limitations of Existing Approaches.

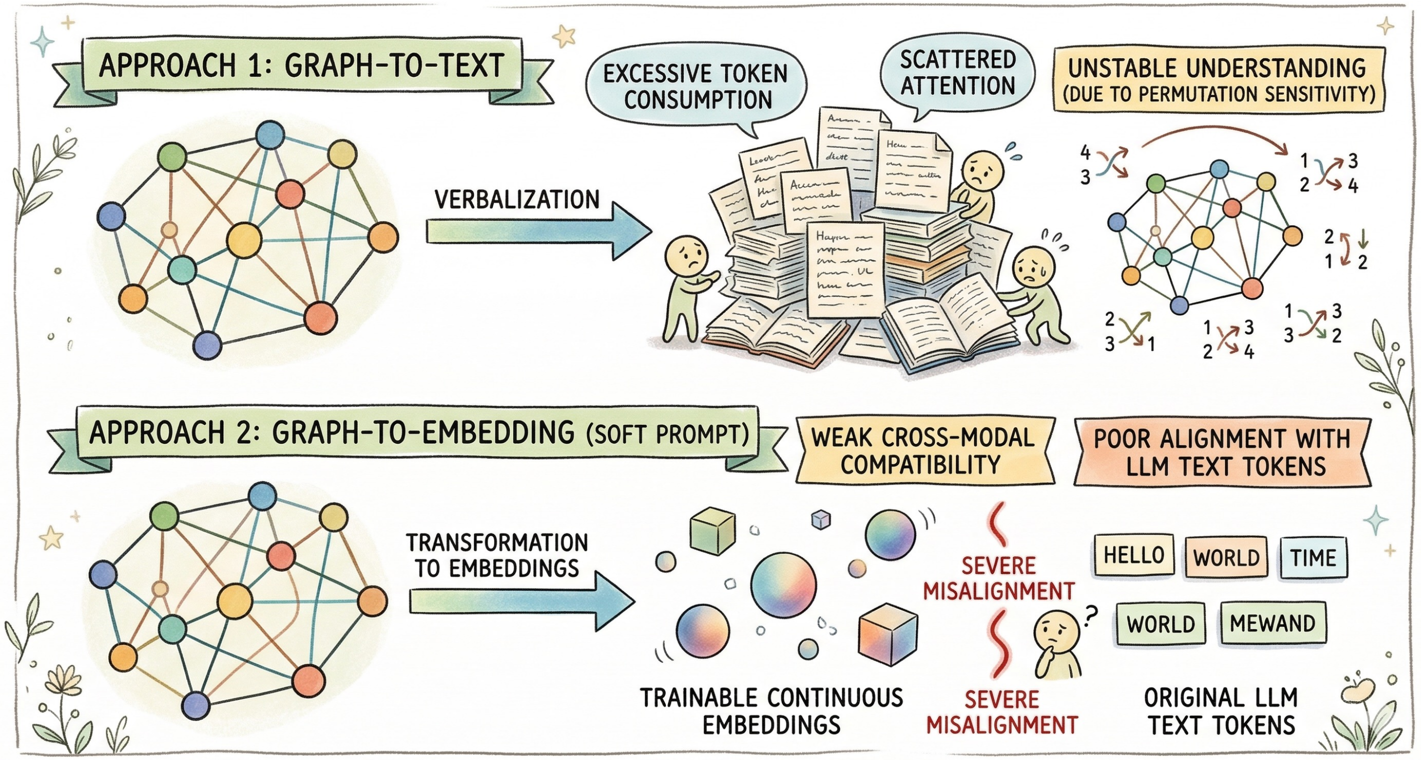
- ➢ **Graph-to-Text**: excessive token consumption, scattered attention and permutation sensitivity.(not concise)
- ➢ **Graph-to-Embedding**: severe misalignment, weak cross-modal compatibility. (not accurate)



Introduction: LLMs for Graph Understanding

■ Limitations of Existing Approaches.

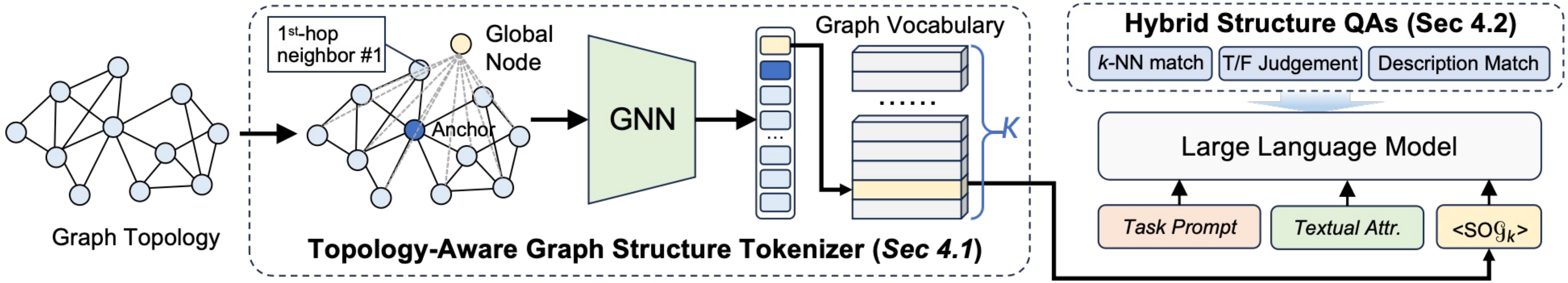
- **Graph-to-Text**: excessive token consumption, scattered attention and permutation sensitivity.(not concise)
- **Graph-to-Embedding**: severe misalignment, weak cross-modal compatibility. (not accurate)



How to concisely and accurately feed graph topology information into an LLM?

Methodology

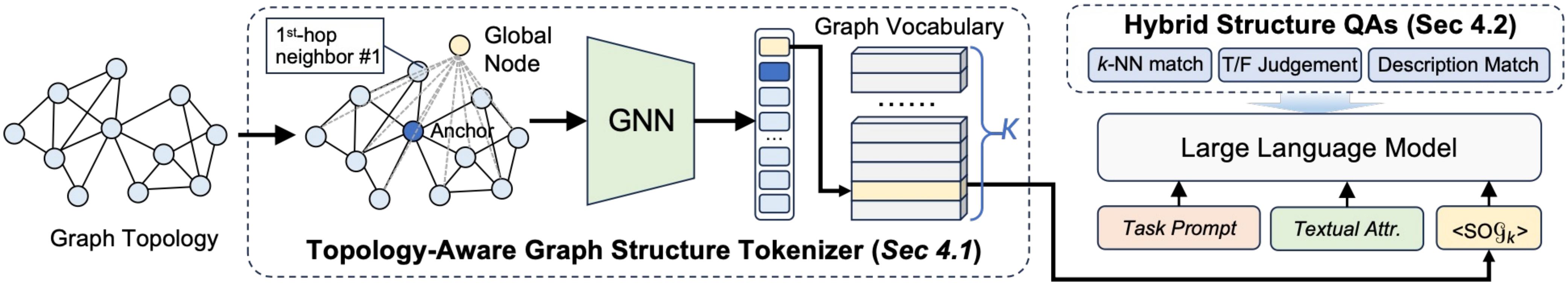
■ Framework of $< SO\mathcal{G}_k >$



Methodology

■ Framework of $< \mathcal{SOG}_k >$

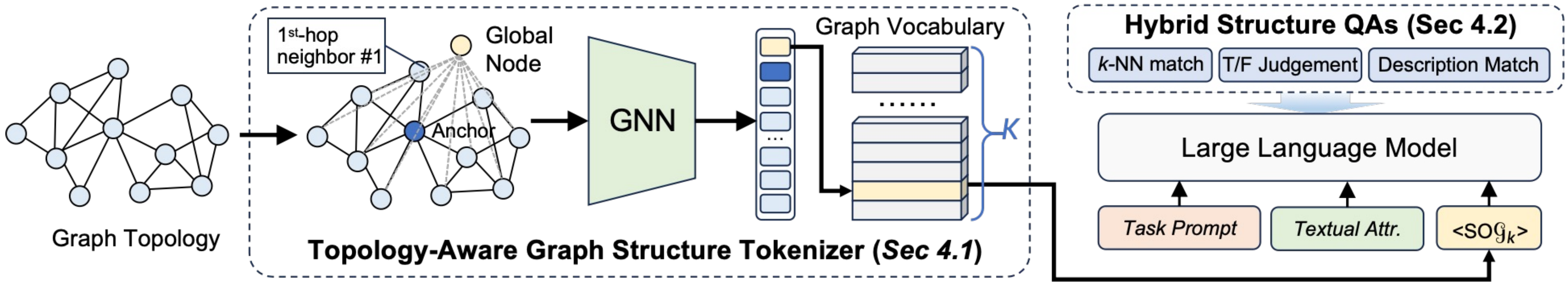
- ➢ **Topology-Aware Graph Tokenizer**:
  - ➢ Map each topology to **only one structural token** $< \mathcal{SOG}_k >$ to provide structural information to LLMs (**concise**)



Methodology

■ **Framework of $< \boldsymbol{SOG_k} >$**

- ➢ **Topology-Aware Graph Tokenizer**:
  - ➢ Map each topology to **only one structural token** $< \boldsymbol{SOG_k} >$ to provide structural information to LLMs (**concise**)
- ➢ **Hybrid Structure QAs Alignment**:
  - ➢ Design topology-based hybrid QAs instruction tuning to align the embedding space of textual and structural tokens (**accurate**)



Methodology

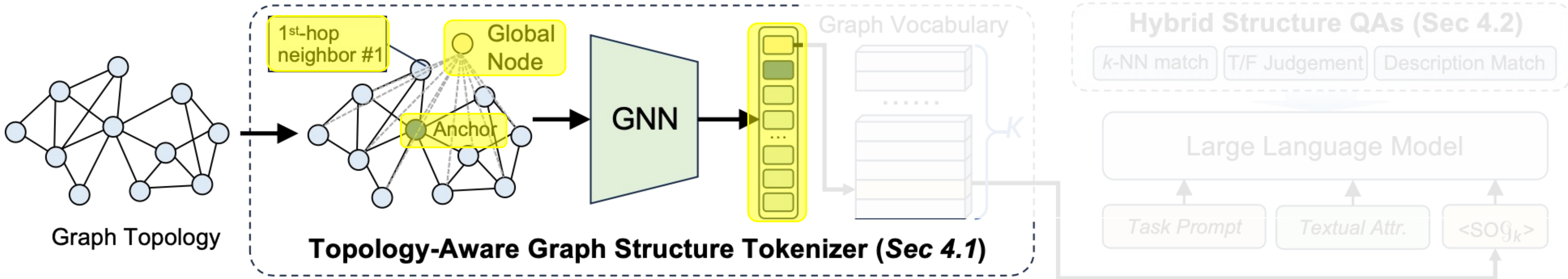
■ **Topology-Aware Graph Structural Tokenizer:**

- Each graph gets an explicit topology-aware structural descriptor < $\boldsymbol{SO\mathcal{G}_k}$ > and nodes are pooled by a global node.

For each graph we extract its topology $\mathrm{G} = (V, E)$ and choose the most ”important” node as **anchor**. Based on **anchor node** we provide each node a spatial coordinate to reinitialize node features. Finally, we add a virtual **global node** to summarize graph-level information to get $X_s = [\, x_s^1; x_s^2; \ldots; x_s^{|V|}; x_s^{\text{global}} \,] \in \mathbb{R}^{(|V|+1)\times d_s}$.

- GNN turns explicit structural attributes into topology-aware latent features

$$H_s = f_G(X_s),\, H_s \in \mathbb{R}^{(|V|+1)\times d}$$



Methodology

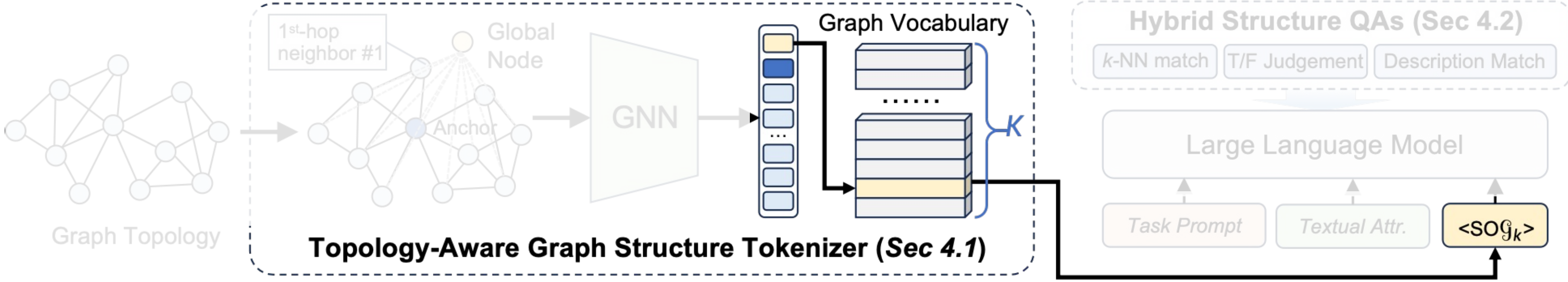
■ Topology-Aware Graph Structural Tokenizer:

- Continuous topology features are discretized into LLM-friendly structural vocabulary.

We introduce a learnable vocabulary $C = \{c_1, c_2, \ldots c_k\}$ and retrieve the nearest vocabulary entry based on Euclidean distance:

$$k = \arg\min_j \left\| h_s^i - c_j \right\|_2, \quad c_j \in \mathcal{C} = \{c_1, c_2, \ldots, c_K\}.$$

Finally, code selection of the global node is chosen as structural token for the graph.



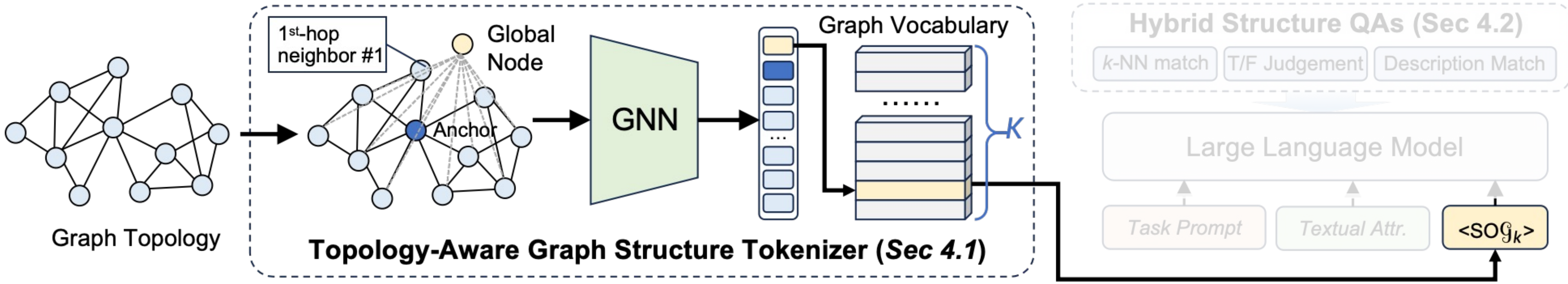
Methodology

■ Topology-Aware Graph Structural Tokenizer:

- Continuous topology features are discretized into LLM-friendly structural vocabulary.

Use three losses jointly to guide codebook selection:

Reconstruction loss: preserve graph topology by reconstructing adjacency $\hat{X} = f_q(\mathbf{z}_e(H_s))$ $\hat{A} = \hat{X}\hat{X}^T$
Codebook update loss: pull codewords toward encoded features
Commitment loss: force encoded features to stay close to selected codewords



Methodology

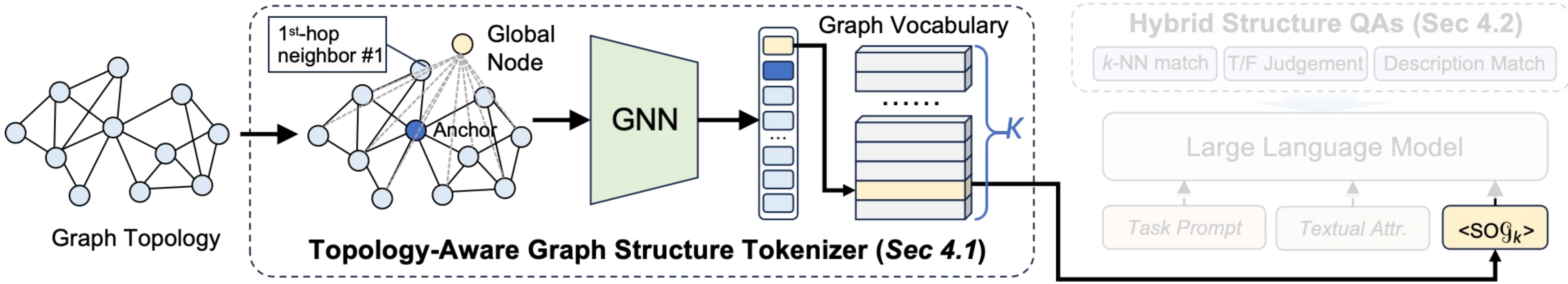
■ Topology-Aware Graph Structural Tokenizer:

- Continuous topology features are discretized into LLM-friendly structural vocabulary.

Use three losses jointly to guide codebook selection:

Reconstruction loss: preserve graph topology by reconstructing adjacency $\hat{X} = f_q(\mathbf{z}_e(H_s))$ $\hat{A} = \hat{X}\hat{X}^T$
Codebook update loss: pull codewords toward encoded features
Commitment loss: force encoded features to stay close to selected codewords

$$\mathcal{L} = \underbrace{\left\|A-\hat{A}\right\|_F^2}_{\text{Reconstruction loss}} + \underbrace{\|\mathrm{sg}[H_s]-\mathbf{z}_e(H_s)\|_2^2}_{\text{Update loss}} + \beta\underbrace{\|H_s - \mathrm{sg}[\mathbf{z}_e(H_s)]\|_2^2}_{\text{Commitment loss}},$$



Methodology

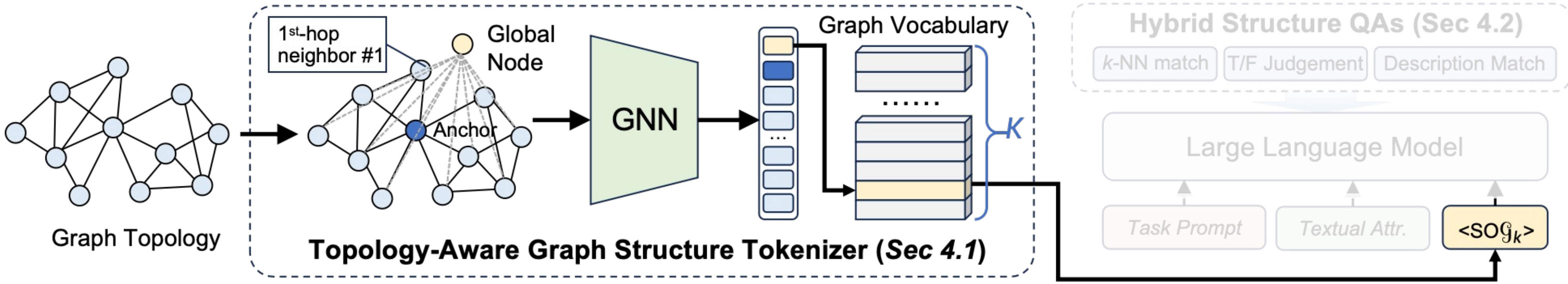
■ Topology-Aware Graph Structural Tokenizer:

- Continuous topology features are discretized into LLM-friendly structural vocabulary.

Use three losses jointly to guide codebook selection:

Reconstruction loss: preserve graph topology by reconstructing adjacency $\hat{X} = f_q(\mathbf{z}_e(H_s)) \quad \hat{A} = \hat{X}\hat{X}^T$
Codebook update loss: pull codewords toward encoded features
Commitment loss: force encoded features to stay close to selected codewords

$$\mathcal{L} = \underbrace{\left\| A - \hat{A} \right\|_F^2}_{\text{Reconstruction loss}} + \underbrace{\|\mathrm{sg}[H_s] - \mathbf{z}_e(H_s)\|_2^2}_{\text{Update loss}} + \beta \underbrace{\|H_s - \mathrm{sg}[\mathbf{z}_e(H_s)]\|_2^2}_{\text{Commitment loss}},$$



Methodology

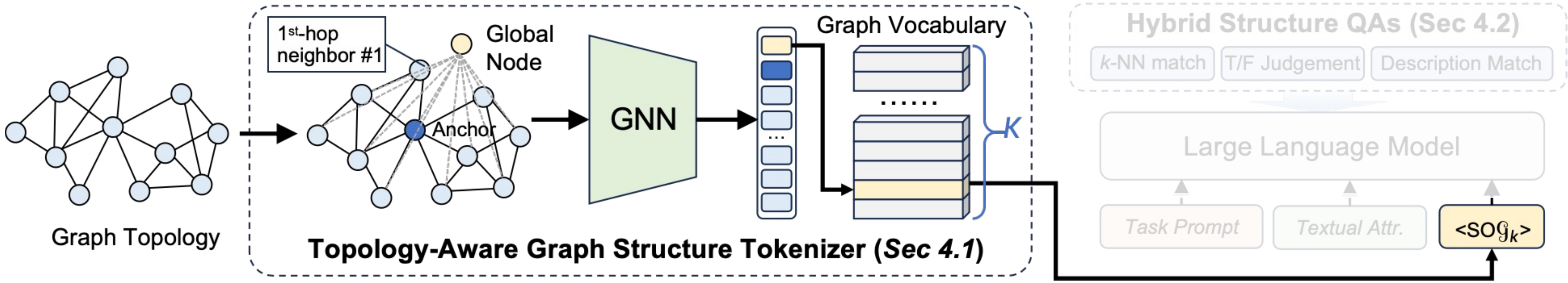
■ Topology-Aware Graph Structural Tokenizer:

- Continuous topology features are discretized into LLM-friendly structural vocabulary.

Use three losses jointly to guide codebook selection:

Reconstruction loss: preserve graph topology by reconstructing adjacency $\hat{X} = f_q(\mathbf{z}_e(H_s)) \quad \hat{A} = \hat{X}\hat{X}^T$
Codebook update loss: pull codewords toward encoded features
Commitment loss: force encoded features to stay close to selected codewords

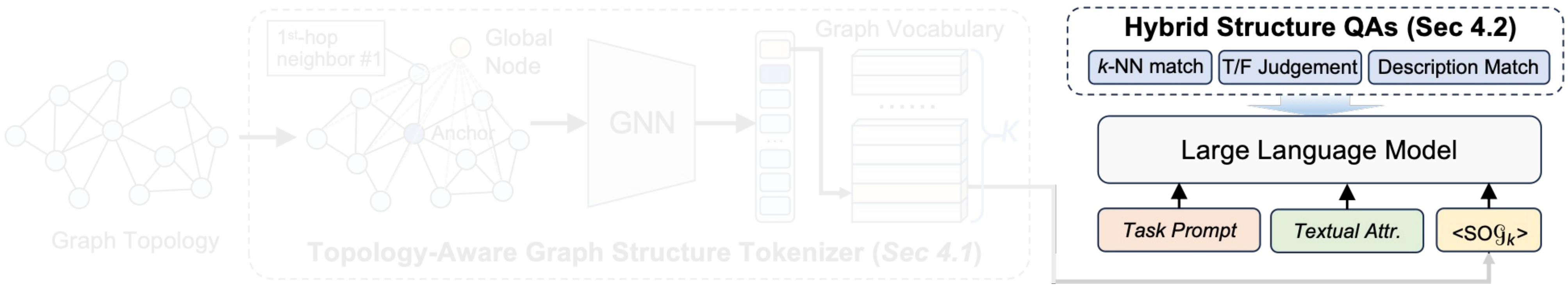
$$\mathcal{L} = \underbrace{\left\| A - \hat{A} \right\|_F^2}_{\text{Reconstruction loss}} + \underbrace{\|\mathrm{sg}[H_s] - \mathbf{z}_e(H_s)\|_2^2}_{\text{Update loss}} + \beta \underbrace{\|H_s - \mathrm{sg}[\mathbf{z}_e(H_s)]\|_2^2}_{\text{Commitment loss}},$$



Methodology

■ **Token Alignment with Hybrid Structure QAs:**

Align newly added structural tokens with native LLM text tokens for unified understanding.

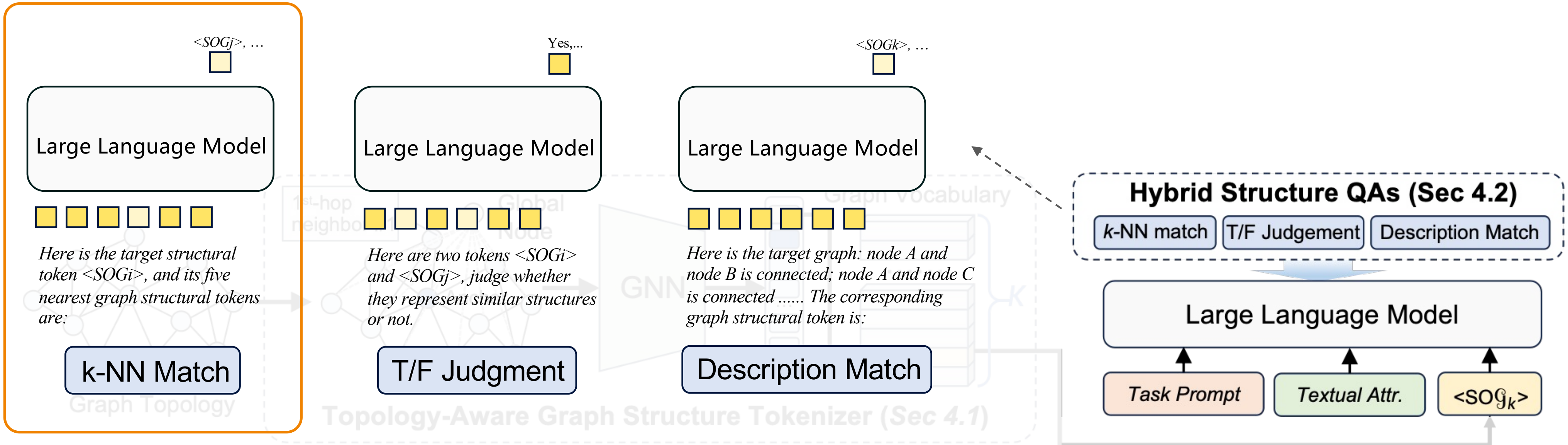


Methodology

■ Token Alignment with Hybrid Structure QAs:

Align newly added structural tokens with native LLM text tokens for unified understanding.

k-NN-Neighbor Matching: Predict top-k nearest structural tokens for a target token. -> encourages local agreement.

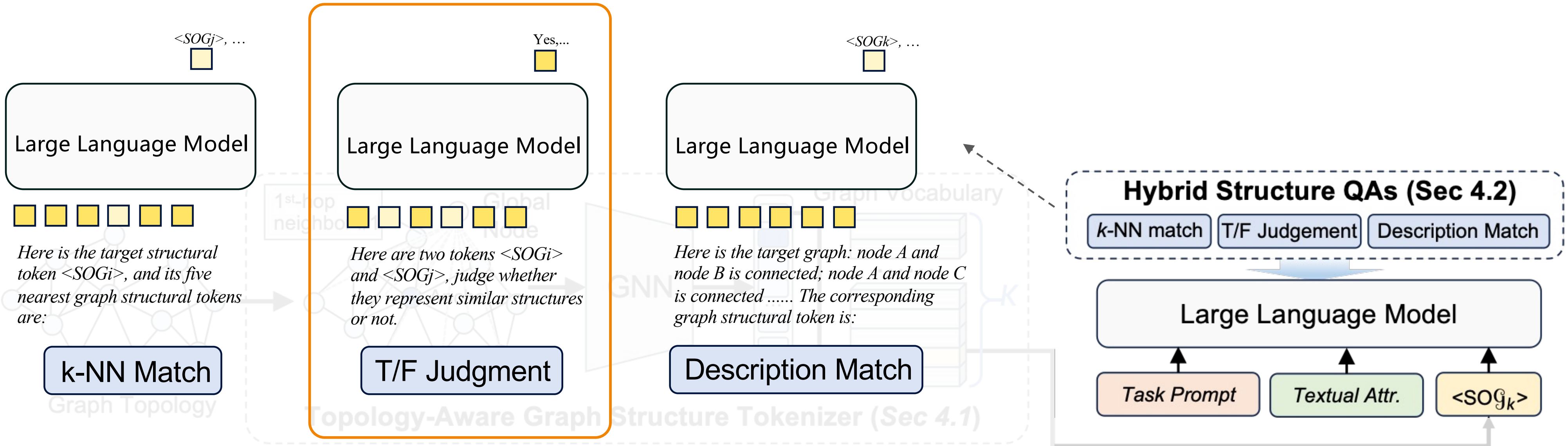


Methodology

■ Token Alignment with Hybrid Structure QAs:

Align newly added structural tokens with native LLM text tokens for unified understanding.

k-NN-Neighbor Matching: Predict top-k nearest structural tokens for a target token. -> encourages local agreement.
T/F Similarity Judgment: Predict whether two structural tokens represent similar structures.-> clarifies similarity boundary.

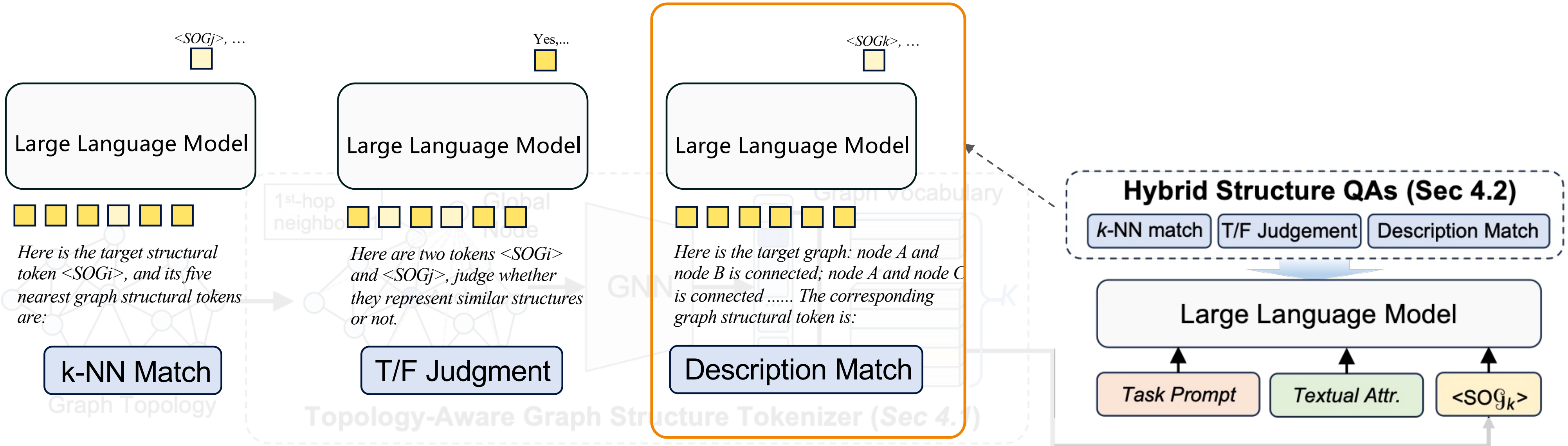


Methodology

■ **Token Alignment with Hybrid Structure QAs:**

Align newly added structural tokens with native LLM text tokens for unified understanding.

k-NN-Neighbor Matching: Predict top-k nearest structural tokens for a target token. -> encourages local agreement.
T/F Similarity Judgment: Predict whether two structural tokens represent similar structures.-> clarifies similarity boundary.
Description-Token Matching: Match a textual structural description to its corresponding token.-> bridges text and structural understanding.



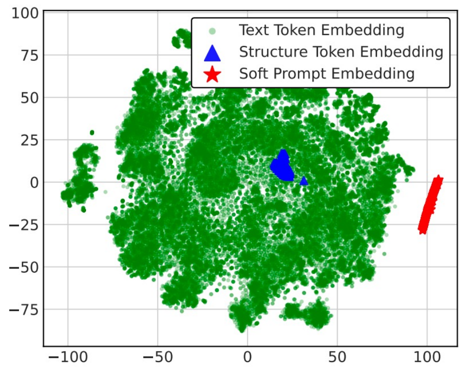
Experiment

■ Performance Comparison

➢ $< \boldsymbol{SOG_k} >$ achieves the **highest average performance** across all five graph benchmarks while effectively extending to node-level tasks, and effectively aligning with textual tokens in LLMs.

LLM Backbone	Method	BBBP (↑)	Tox21 (↑)	ClinTox (↑)	HIV (↑)	BACE (↑)
GPT-4	Zero-shot	61.5	55.2	51.6	65.9	62.5
GPT-4o	Zero-shot	57.0	36.0	51.8	54.7	38.5
Deepseek-R1	Zero-shot	63.6	60.1	48.2	50.5	52.7
LLaMA3-70B	Zero-shot	60.1	44.9	48.8	58.3	50.9
Galactica-120B	Zero-shot	66.1	68.9	82.6	74.5	61.7
Non-LLM Molecule Approach	GPF-AttrMasking	58.8	66.0	-	71.3	62.2
	GPF-GCL	56.5	63.6	-	49.3	52.1
	GIMLET	59.4	61.2	-	66.2	69.6
	MolCA-S	62.5	66.6	-	69.0	62.8
	MolCA-GS	63.6	68.5	-	72.7	63.9
LLaMA3-3B	Zero-shot	50.2 ± 3.4	48.5 ± 2.7	50.2 ± 2.5	51.8 ± 0.8	52.7 ± 1.8
	Few-shot (Random)	46.5 ± 3.8	49.9 ± 10.2	57.8 ± 6.2	49.7 ± 16.7	68.2 ± 17.5
	Few-shot (Morgan Sim)	49.7 ± 2.5	51.0 ± 8.7	53.3 ± 10.0	42.0 ± 14.7	49.7 ± 0.6
	LoRA SFT	60.2 ± 1.2	50.8 ± 0.8	63.8 ± 2.8	51.5 ± 0.7	50.5 ± 2.9
	Soft Prompt	27.9 ± 1.7	39.9 ± 1.9	35.0 ± 2.0	33.6 ± 0.5	28.5 ± 0.9
	<SO$\mathcal{G}_k$> (Ours)	**76.9 ± 3.1**	**83.4 ± 3.3**	85.5 ± 3.7	75.7 ± 1.6	63.3 ± 4.2
LLaMA2-7B	Zero-shot	50.0 ± 0.0	48.4 ± 3.3	44.0 ± 2.6	50.0 ± 0.0	49.9 ± 0.4
	Few-shot (Random)	49.4 ± 0.6	49.3 ± 2.2	50.3 ± 1.7	46.9 ± 0.8	47.8 ± 1.8
	Few-shot (Morgan Sim)	51.3 ± 0.8	51.1 ± 2.5	56.8 ± 6.3	50.4 ± 0.7	53.2 ± 1.5
	LoRA SFT	55.0 ± 0.8	60.6 ± 3.4	59.3 ± 3.7	78.1 ± 2.2	61.7 ± 2.7
	Soft Prompt	53.7 ± 0.8	49.8 ± 1.5	49.7 ± 4.7	44.9 ± 0.2	49.4 ± 2.3
	<SO$\mathcal{G}_k$> (Ours)	66.4 ± 2.7	72.4 ± 3.1	**94.3 ± 0.1**	**83.2 ± 1.9**	**98.4 ± 0.8**

Model	Cora		Pubmed	
	Acc (↑)	F1 (↑)	Acc (↑)	F1 (↑)
DGI	17.50	12.44	44.88	38.72
GraphMAE	27.08	23.66	22.03	15.65
LLaMA3 (70B)	67.99	68.05	77.00	64.18
GPT-3.5-turbo	65.67	63.22	75.99	69.90
GPT-4o	68.62	68.49	77.96	71.79
DeepSeek-chat	65.62	65.77	79.23	74.30
Emb w/ NA	63.59	58.23	74.66	73.15
OFA	23.11	23.30	46.60	35.04
ZEROG	62.52	57.53	79.08	77.94
GraphGPT	24.90	7.98	39.85	20.07
Ours (LLaMA3-3B)	**91.58**	**78.62**	**97.46**	85.50
Ours (LLaMA2-7B)	88.80	71.28	96.27	**89.64**



Experiment

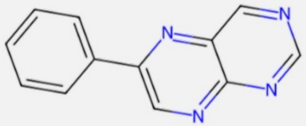
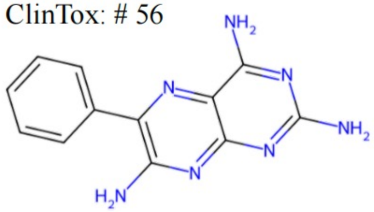
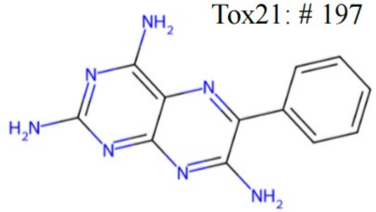
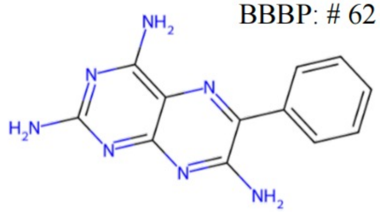
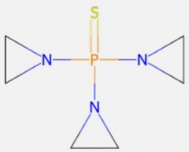
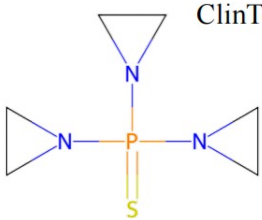
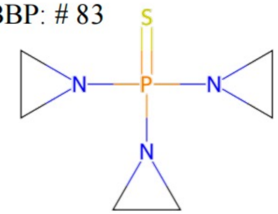
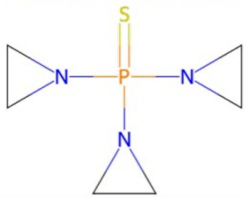
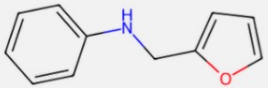
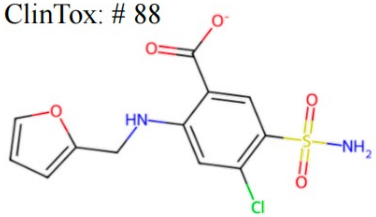
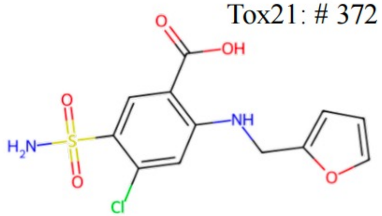
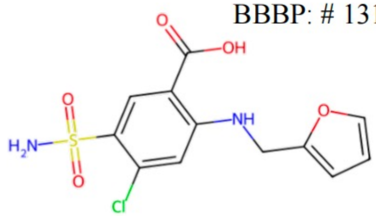
■ Case Study, Ablation Study and Further Eexperiments

- Case study demonstrate consistent alignment, high selectivity and low redundancy of structural tokens.

Experiment

■ Case Study, Ablation Study and Further Eexperiments

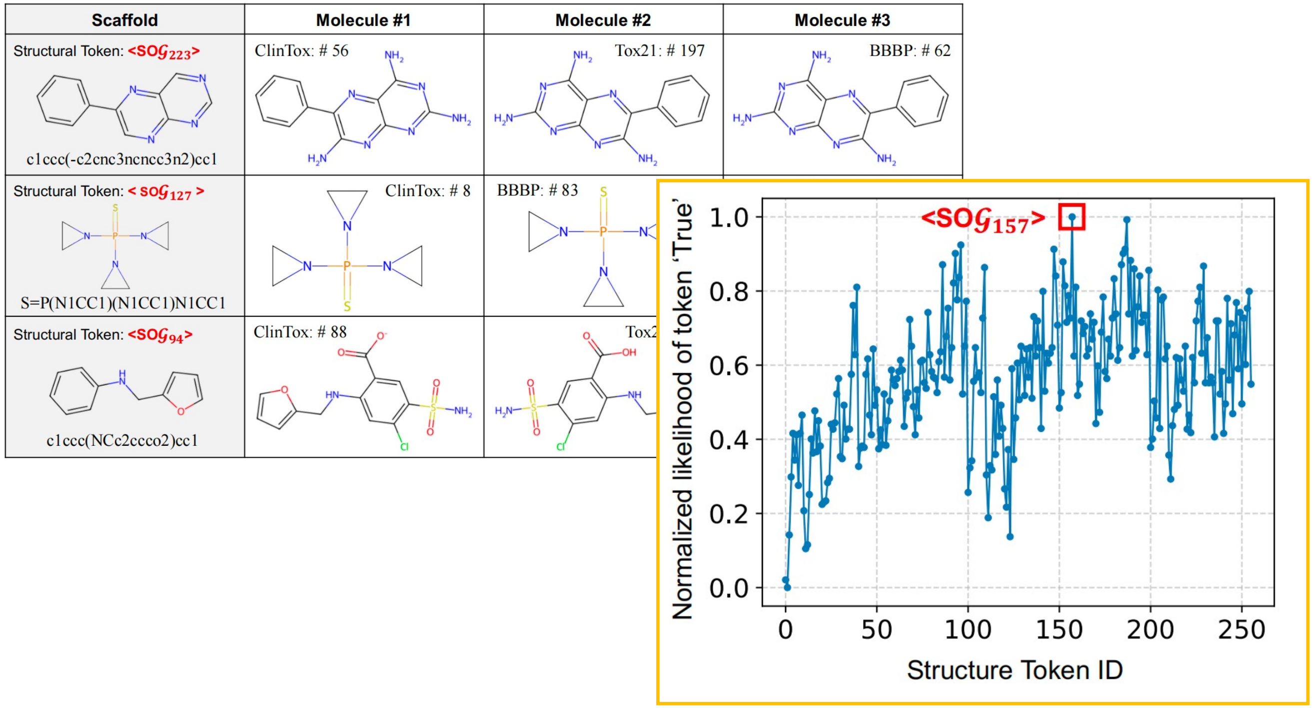
➢ Case study demonstrate consistent alignment, high selectivity and low redundancy of structural tokens.

Scaffold	Molecule #1	Molecule #2	Molecule #3
Structural Token: <SOG_{223}>  c1ccc(-c2cnc3ncncc3n2)cc1	ClinTox: # 56 	Tox21: # 197 	BBBP: # 62 
Structural Token: < SOG_{127} >  S=P(N1CC1)(N1CC1)N1CC1	ClinTox: # 8 	BBBP: # 83 	HIV: # 4021, Tox21: #83 
Structural Token: <SOG_{94}>  c1ccc(NCc2ccco2)cc1	ClinTox: # 88 	Tox21: # 372 	BBBP: # 131 

Experiment

■ Case Study, Ablation Study and Further Eexperiments

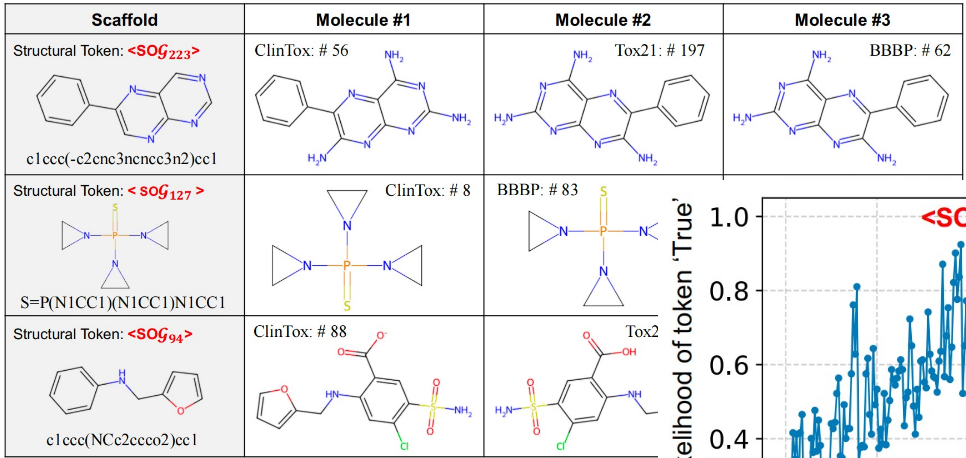
➢ Case study demonstrate consistent alignment, high selectivity and low redundancy of structural tokens.

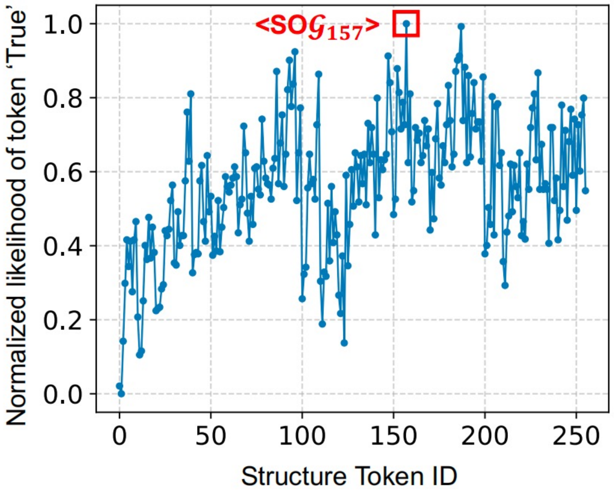


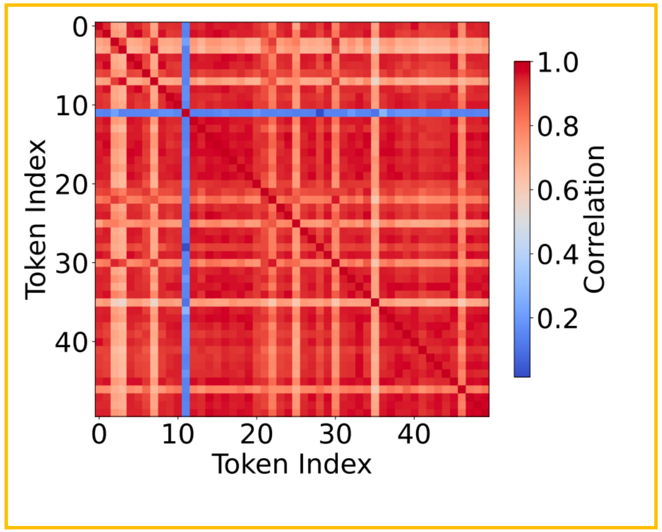
Experiment

■ Case Study, Ablation Study and Further Eexperiments

➢ Case study demonstrate consistent alignment, high selectivity and low redundancy of structural tokens.





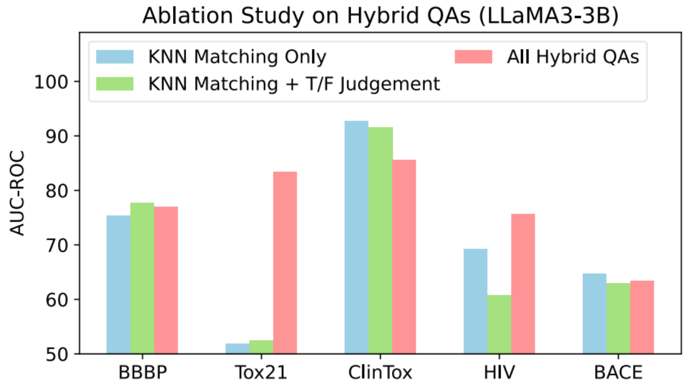


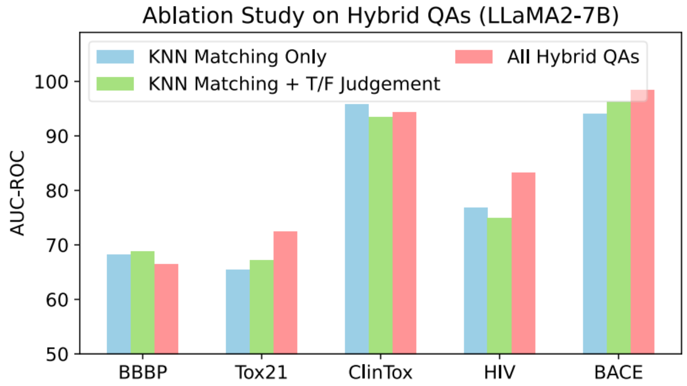
Experiment

■ Case Study, Ablation Study and Further Eexperiments

➢ Ablation studies verify the effectiveness of each module in $< \boldsymbol{SOG_k} >$.

Model	Ablation Study	BBBP	Tox21	ClinTox	HIV	BACE
LLaMA3-3B	(M1a) w/o Structural Token	61.2	72.3	81.8	53.0	53.2
	(M1b) Static Structural Token	60.7	67.8	52.5	54.6	54.3
	(M1c) Random Structural Token	57.3	62.2	52.8	55.3	55.8
	Ours	**76.9**	**83.4**	**85.5**	**75.7**	**63.3**
LLaMA2-7B	(M1a) w/o Structural Token	61.3	66.4	79.5	69.6	94.2
	(M1b) Static Structural Token	64.7	69.1	92.0	81.5	94.2
	(M1c) Random Structural Token	62.1	68.6	88.5	**83.4**	96.7
	Ours	**66.4**	**72.4**	**94.3**	83.2	**98.4**





Conclusion

- **Aligning Graph Topology Perception with LLMs with a Single Token**

- We use a single structural token $< \boldsymbol{SO\mathcal{G}_k} >$ to feed each graph into LLMs and design hybrid structure QAs alignment to enable unified topology understanding.

Thank You!

Thanks
Q&A