Table 16: Comparison of retrieval methods on the WebQSP dataset. This table compares our proposed PCST-based retrieval method with several baseline retrieval strategies, utilizing the llama2-7b-chat model. The baselines include KAPING, which retrieves the top-k triples related to the query, a method that retrieves the top-k nodes and their one-hop neighbors, and a shortest path retrieval approach that selects the top-k nodes and the shortest paths between them. All methods are evaluated using k = 5. The comparison highlights the effectiveness of PCST-based retrieval.

Method	Hit@1
PCST-based retrieval	66.17
top-k triples retrieval (KAPING)	52.64
top-k nodes plus its neighbors	49.82
shortest path retrieval	55.20

Table 17: Comparative Accuracy of *G-Retriever* and Fixed LLMs on the WebQSP Dataset. This table compares the performance of G-Retriever to fixed LLMs (llama2-7b-chat and GPT-4o). In the fixed LLM setup, PCST retrieval was used to extract relevant subgraphs, which were then converted into text and input into the LLMs. The results show that G-Retriever outperforms the fixed LLM baselines.

Method	Hit@1
llama2-7b-chat	66.17
GPT-4o	67.87
G-Retriever	70.49

Table 18: This table compares the inference times and accuracy of *G-Retriever* with and without retrieval. The variant without retrieval skips the additional steps of node and edge retrieval and subgraph construction using the PCST algorithm. Despite these additional steps, *G-Retriever* achieves significantly faster inference times and higher accuracy.

	Time in minutes	Accuracy with Hit@1
G-Retriever	9.55	70.49
G-Retriever w/o retrieval	21.01	63.84

Table 19: Retrieval success rate comparison between our PCST-based subgraph retrieval method and the top-k triple retrieval method from KAPING on the WebQSP dataset. A retrieval is considered successful if the correct label is included within the retrieved subgraph.

Method	Hit@1
top-k triple retrieval (KAPING)	60.81%.
PCST-based subgraph retrieval	67.87%

Table 20: Impact of node and edge text attributes on G-Retriever performance across different stages: Retrieval, GNN Input, and LLM Input. The table shows results with and without these attributes to evaluate their contribution.

	without Node	without Edge
Retrieval	66.58	58.37
GNN Input	68.85	67.87
LLM Input	56.32	68.24