Transactions on Computational and Scientific Methods | Vo. 4, No. 11, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

# Bootstrapped Structural Prompting for Analogical Reasoning in Pretrained Language Models

**Yue Xing**
University of Pennsylvania, Philadelphia, USA
jasmineyxing@gmail.com

**Abstract:** This paper proposes an analogical reasoning method based on structured prompt graphs and bootstrapped path optimization. The goal is to improve the reasoning ability of large language models in complex structural transfer tasks. The method first constructs an analogical prompt graph that transforms input analogy pairs into explicit structural representations. This captures relational paths and topological dependencies between concepts. On this basis, a bootstrapped path optimization mechanism is introduced. It scores initial reasoning paths based on semantics and filters them for structural consistency. The prompt graph is then dynamically updated to strengthen the transfer expression of analogical structures. The model performs multi-round path reconstruction to achieve stable modeling and iterative refinement of analogical structures. This enhances the robustness and consistency of the language model in scenarios involving multi-hop reasoning, few-shot analogy, and path uncertainty. The experiments systematically evaluate the proposed method under different path depths, task types, and sample sparsity levels. The evaluation focuses on accuracy, path consistency, and structural alignment. The results show that the proposed method outperforms existing mainstream models across all metrics. It demonstrates significant advantages in structural reasoning, stability, and cross-task transfer. These findings validate the modeling value of structured prompting and bootstrapped mechanisms in analogical reasoning tasks.

**Keywords:** Analogical reasoning; structural perception cues; path optimization; structural consistency

## 1. Introduction

With the remarkable progress of large language models in natural language understanding and generation tasks, enhancing their analogical reasoning ability has become a major challenge in current artificial intelligence research. Analogical reasoning is a core component of human cognition. It is widely applied in mathematics, language, scientific discovery, and other domains[1,2]. At its core, analogical reasoning involves transferring known relational structures between concept pairs to novel inference scenarios. Although existing language models demonstrate strong capabilities in capturing and generating semantics, they often rely on pattern matching when facing highly structured or cross-domain analogical reasoning tasks. They lack deep structural understanding and relational transfer abilities. This limitation constrains the application of language models in higher-order reasoning tasks[3].

Current mainstream training paradigms for language models are based on large-scale self-supervised learning. Models learn language patterns and contextual associations from massive corpora, which gives them a certain level of reasoning capability. However, this language-continuity-based training objective often leads to fitting surface-level patterns rather than abstracting structural relationships. In analogical reasoning, models tend to recognize semantic similarity but fail to grasp deeper structural analogical patterns. These include relation mapping, structural alignment, and recursive analogy. This reveals the expressive and controllability bottlenecks of traditional prompting in analogical scenarios. To enable language models to reason more like humans, it is crucial to design structurally-aware prompting mechanisms. These should guide models to construct and transfer structural patterns during generation and reasoning processes.

Structured Prompt Graphs have emerged as a novel prompting approach that provides a more systematic way to organize and guide the reasoning process of language models. Unlike traditional linear prompts, structured prompt graphs explicitly represent concept nodes and their relations in a graph structure. This allows models to recognize and leverage complex relational dependencies and graph-based patterns. In analogical reasoning, these graph structures can represent equivalence between concept pairs. They can also encode multi-level analogical paths within the graph. This builds a transferable structural foundation. With structured prompt graphs, the model processes input in an explicit relational space rather than a linear sentence space. This offers a natural modeling space for structural alignment and relational transfer in analogical reasoning[4].

A core challenge in analogical reasoning is how to induce generalizable structural relations from limited examples and transfer them to new contexts. To address this, a bootstrapping mechanism is introduced to improve generalization and robustness in the reasoning process. Bootstrapping emphasizes the model's ability to extract its own reasoning patterns through repeated analogical attempts and refine them via feedback. When combined with structured prompt graphs, bootstrapping supports pattern extraction at both the example and graph levels. It enables reconstruction and selection of analogical paths within the graph. This achieves joint optimization of structure and semantics. The combined approach shows strong potential in improving the model's perception, abstraction, and transfer of complex analogical structures. It lays the foundation for building reasoning systems with structural cognition[5].

Research on analogical reasoning in large language models based on structured prompt graphs and bootstrapping not only expands the theoretical modeling of structural knowledge and reasoning processes but also offers new paradigms and methodological support for high-level reasoning tasks. This direction helps bridge the gap between language understanding and logical reasoning in current models[6,7]. It promotes a paradigm shift from language-driven generation to structure-driven reasoning. In the development of future general artificial intelligence systems, structural reasoning abilities such as analogy, induction, and transfer will become essential components. Therefore, this research has long-term and significant academic and practical value.

## 2. Related work

### 2.1 Large Language Model

In recent years, large language models have become a central pillar in natural language processing. Trained on large-scale corpora with self-supervised objectives, these deep neural architectures exhibit strong cross-task transfer capabilities[8]. They have reached or surpassed human-level performance in traditional tasks such as text generation, question answering, summarization, and translation. With the exponential growth in model size, their abilities in semantic understanding, language interaction, and reasoning have also improved significantly. These models are typically built on the Transformer architecture. They learn the statistical structure of language through autoregressive or autoencoding mechanisms. However, while this data-driven learning approach captures a vast amount of vocabulary and contextual associations, its reasoning ability remains largely dependent on pattern reuse and surface-level semantic similarity[9,10,11]. As a result, it shows clear limitations when dealing with tasks that require stronger structural understanding or complex logic.

In recent studies, many approaches have been proposed to enhance the reasoning abilities of language models. These include Chain-of-Thought prompting, Retrieval-Augmented Generation, and tool use. Such methods have improved model interpretability and task performance to some extent, especially in mathematical reasoning and question answering. However, from a more fundamental perspective, these approaches rely on external components or human-designed templates[12]. They do not fundamentally address the internal weakness of language models in structural awareness. Traditional prompts are usually presented as flat text sequences. They are not able to encode complex semantic dependencies or hierarchical relations. They also fail to offer explicit structural guidance to the model. This limitation in representation reduces the expressive capacity of models in high-level reasoning scenarios such as analogical thinking and inductive summarization.

To overcome these limitations, researchers have begun introducing structural modeling concepts into prompt design. They aim to compensate for the model's lack of internal structural understanding by incorporating explicit structural information into its inputs. By using structured prompts in the form of graphs, tree representations, or code snippets, the model can learn more abstract representations from these explicit relations. This is particularly effective in complex tasks involving multiple entities and relationships. Structured inputs not only enhance the model's perception of such relations but also significantly improve its ability to model logical connections among concepts in context. This approach also lays a methodological foundation for more controllable generation, traceable reasoning, and robust analogical induction. Structure-guided language models are advancing the shift from data-driven weak reasoning toward structure-supported strong reasoning. This transition is building both the theoretical and practical foundation for more capable and general-purpose language intelligence systems[13].

### 2.1 Structured Prompt Graph

Traditional prompting methods typically use linear text sequences as input to language models. This approach is easy to implement and broadly compatible. However, it has inherent limitations in expressing complex relations or structural constraints[14,15]. Linear prompts cannot clearly represent dependency hierarchies, logical paths, or topological structures between concepts[16]. As a result, models often fall into shallow pattern matching when handling tasks that require structural alignment, inductive transfer, or multi-hop reasoning. To address the structural deficiency in prompt inputs, research on structured prompting has gained increasing attention. The core idea is to explicitly construct graph structures with semantic relations, organizing language elements as nodes and edges. This allows the language model to understand and utilize the topological associations embedded in the input, leading to more effective reasoning in structurally demanding tasks[17].

Structured prompt graphs not only change how prompts are represented but also expand the pathways through which models process information. By representing entities, concepts, and propositions as nodes, and modeling logical, causal, equivalent, or sequential relations as edges, prompt graphs create a non-linear and traversable structure for the model. This representation helps the model select reasoning paths more accurately during generation[18]. It also supports semantic alignment and preserves contextual coherence. In tasks such as chain reasoning, graph-based question answering, and analogical inference, structured prompt graphs clearly present inference chains and dependency structures. This helps the model learn the structural transitions between reasoning steps in an explicit manner. Compared with traditional prompts, prompt graphs show greater potential in structure compression, cross-task generalization, and intermediate representation learning.

Structured prompt graphs also offer richer control mechanisms for training and inference. During training, they can serve as intermediate supervision signals that guide the model in learning structural relationships more precisely. During inference, the graph structure can constrain the generation space, preventing semantic drift or logical leaps. When combined with bootstrapping mechanisms, the prompt graph can be expanded and refined across multiple reasoning rounds. This helps form stable and efficient analogical mapping structures. This method improves the model's understanding of complex input structures. It also supports controllable generation and fine-grained reasoning through path-dependent modeling. As structure-aware modeling becomes more integrated with language generation tasks, structured prompt graphs have emerged as a key research direction for enhancing reasoning performance. They are evolving into a general and scalable prompting paradigm.

## 3. Method

This study proposes a bootstrapped analogical reasoning method for language models based on structured prompt graphs. The goal is to enhance large language models in structural transfer and analogical construction. First, an Analogical Prompt Graph Generation (APGG) mechanism is introduced. It encodes concept pairs from analogical relations into a graph structure. This explicitly represents structural mapping paths between source and target domains. The graph provides a relationally traversable input space, which strengthens the model's ability to perceive and model structural alignment. Second, a Bootstrapped Path Refinement (BPR) module is integrated. During inference, it constructs multiple candidate reasoning paths based on the model's own outputs. Through path selection and a structural consistency feedback mechanism, the analogical mappings in the prompt graph are iteratively updated. This gradually improves the model's stability and generalization in analogical reasoning under complex structures. Overall, the method establishes a closed-loop framework of structure-guided input and self-supervised optimization. It offers a new approach to enhancing high-level cognitive reasoning in language models.

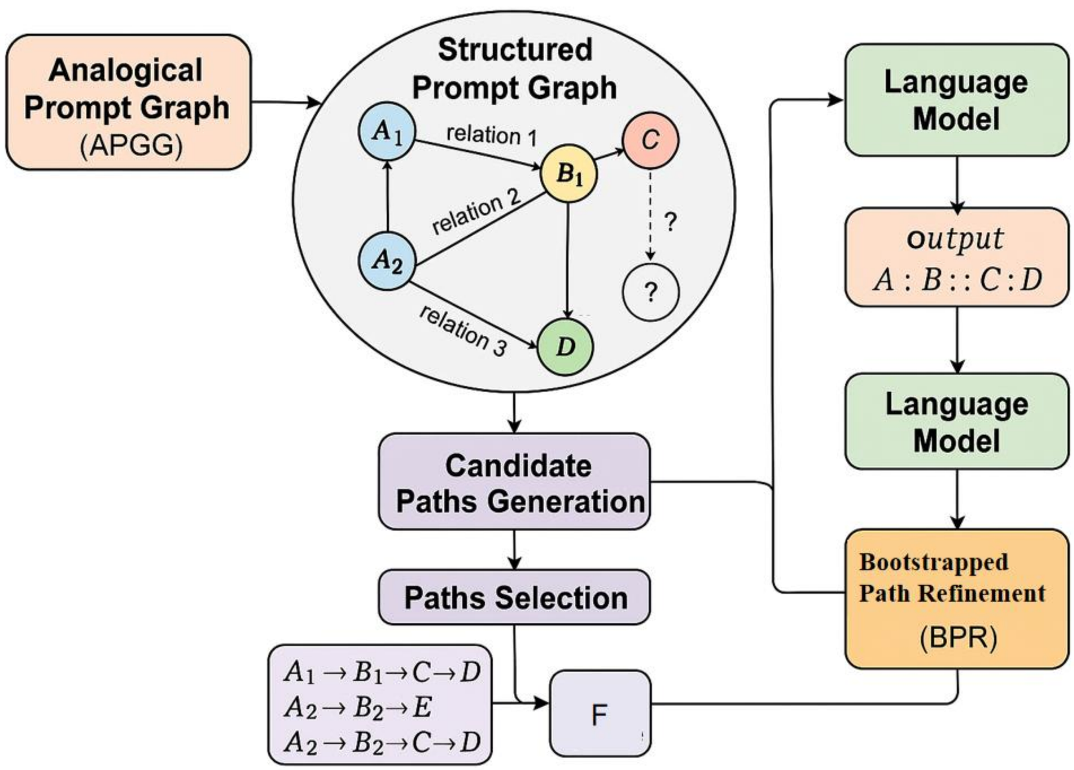


**Figure 1.** Overall model architecture diagram

### 3.1 Analogical Prompt Graph Generation

In this method, the generation process of the analogy hint graph is modeled as a structural mapping task, which aims to transform a set of input analogy pairs into a structurally clear and interpretable graph. The purpose is to capture both the relationship mapping between concept pairs and the underlying structural alignment across different domains or contexts. This process begins by identifying the core entities involved in the analogy and representing them as nodes. The semantic or relational links between these entities are then encoded as edges, forming a graph that reflects the analogical structure in an explicit and organized manner. The resulting graph not only preserves the pairwise analogical relations but also reveals higher-order patterns such as topological dependencies and relational hierarchies. This structured representation enables the model to reason over analogy paths more effectively by leveraging the graph's internal organization. The overall module architecture that supports this transformation process is illustrated in Figure 2.

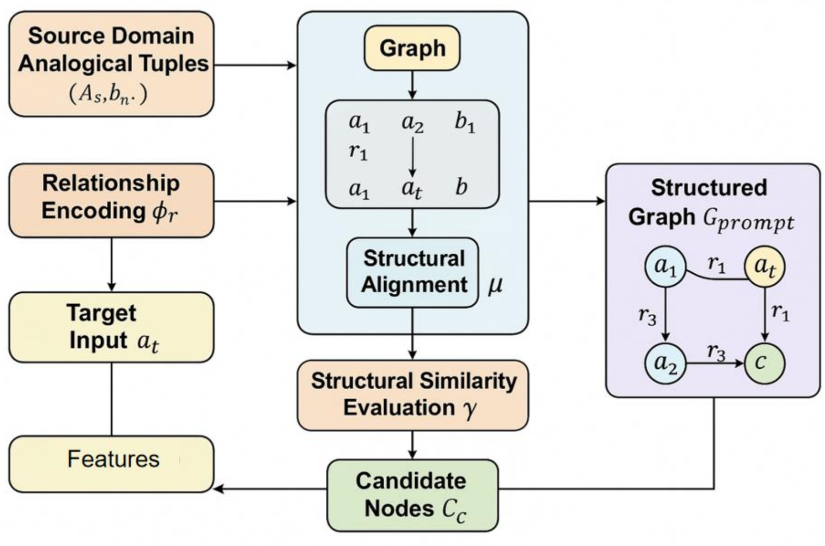


**Figure 2.** APGG module architecture

Let the source domain analogy pair set be $A_s = \{(a_1, b_1), (a_2, b_2), ..., (a_n, b_n)\}$, and the target domain pair to be inferred be $(a_t, ?)$. We first construct an initial graph $G = (V, \varepsilon)$ where the node set V contains all analogy items $a_i, b_i, a_t$, and the edge set $\varepsilon$ represents the corresponding relationship in the source domain.

The construction of the edge is given by a relation encoding function $\phi_r$, which is used to model the analogy relationship between each pair of concepts. The function is defined as:

$$\phi_r(a_i, b_i) = r_i \in R^d$$

Where $r_i$ represents the relationship vector from $a_i$ to $b_i$, which is subsequently used to guide the generation of structural paths. We add edge $(a_i \to b_i; r_i)$ to the graph and define the edge weight of the graph as the norm of the relationship vector or its semantic similarity function.

In order to normalize the structure and constrain the standard configuration, we introduce a structural alignment function $\varphi$ to regularize the paths in the prompt graph:

$$\varphi(P) = Normalize(PathEmbedding(P))$$

Where $P$ represents the path from $a_i$ to $b_i$, and PathEmbedding represents the combined representation of all edges in the path, such as weighted average or sequence aggregation.

For the target input $a_t$, we define a candidate node set $V_c$ and introduce a structural similarity scoring function $\gamma$ to screen the candidate mapping targets that are most similar to the existing structure:

$$\gamma(b,a_t)=\cos(\phi_r(a_i,b_i),\phi_r(a_t,b))$$

Among them, $b \in V_c$ selects the node that satisfies the maximum $\gamma(b,a_t)$ as the initial structure mapping output.

Finally, we constructed a complete analogy prompt diagram:

$$G_{prompt}=GraphConstrust(A_s,a_t,\gamma,\varphi)$$

The structural graph contains the source domain analogy path, edge weight relationship, target node to be inferred, and structural similarity ranking results, which serve as the input for subsequent language model reasoning. Through the above mechanism, the model not only obtains an explicit analogy pair structural representation, but also constructs an analogy migration path in the graph structure, providing structural prior support for subsequent path generation and bootstrap reasoning.

### 3.2 Bootstrapped Path Refinement

In order to enhance the model's reasoning stability and generalization ability for analog structure paths, a bootstrapped path optimization mechanism is further proposed. This mechanism is based on the initial reasoning path and performs path screening and reconstruction in two dimensions: structural consistency and semantic reliability. Its module architecture is shown in Figure 3.

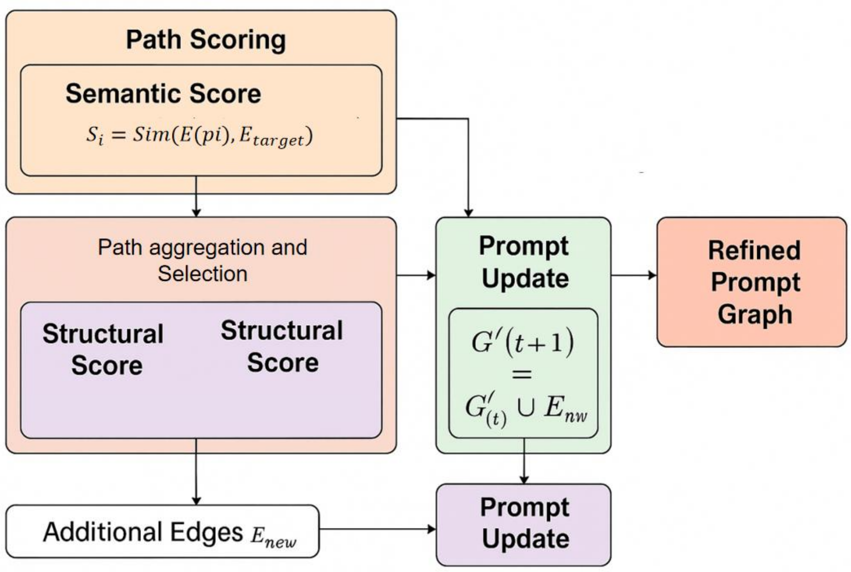


**Figure 3.** BPR module architecture

Let the set of candidate paths generated initially be $P_0=\{p_1,p_2,...,p_k\}$, where each path $p_i$ represents a set of intermediate structural jumps from the source node to the target node. We first calculate the semantic confidence score of each path, which is defined as:

$$s_t^{sem}=Sim(E(p_i),E_{\text{target}})$$

Where $E(p_i)$ represents the overall embedding representation of the path, $E_{\text{target}}$ is the contextual representation of the target class comparison, and Sim represents a scoring function such as cosine similarity or KL divergence.

To further introduce structural consistency constraints, the structural consistency score is defined as:

$$s_t^{sem} = Align(\varphi(p_i), \varphi_{ref})$$

Where $\varphi(p_i)$ is the structural encoding of the path, $\varphi_{ref}$ is the reference structural path extracted from the source domain analogy graph, and Align represents the structural alignment function, which can be achieved through path similarity functions such as Jaccard or subgraph overlap ratio.

After integrating the scores of the two dimensions, we normalize and aggregate the candidate path set and select the top m paths to build the optimized path pool $P_{\text{refine}}$.

$$P_{\text{refine}} = Top_m(\lambda s_i^{sem} + (1-\lambda)s_i^{struct})$$

$\lambda \in [0,1]$ controls the weight balance between semantic and structural scores. This optimized pool will be used in subsequent prompt updates and reasoning cycles.

In the path reconstruction phase, we introduce a bootstrap mechanism to dynamically expand the prompt graph using the results of the previous round of optimization. Let the current structure graph be $G^{(t)}$, and add the edge relationship in the optimized path as a new candidate edge $\varepsilon_{new}$ to the graph, we have:

$$G(t+1) = G(t) \cup \varepsilon_{new}$$

This update mechanism enables the model to gradually accumulate structural knowledge in multiple rounds of reasoning, while enhancing its ability to express complex paths. The refined prompt graph outputted in the end will serve as the input for the subsequent language model to generate analogical relationships, reflecting the dual characteristics of structure-driven and self-supervised convergence.

## 4. Experimental Results

### 4.1 Dataset

This study uses the CLUTRR (Compositional Language Understanding and Text-based Relational Reasoning) dataset as the primary evaluation corpus. This dataset is specifically designed to test language models on relational analogy and reasoning based on text. It contains a large number of synthetic stories generated from relation graphs. Each story describes relationships among several characters. The model is required to infer a hidden relation pair based on the context. CLUTRR is constructed to eliminate the model's reliance on surface-level vocabulary. It encourages the learning of structured and compositional reasoning abilities. Therefore, it is well suited to evaluate the effectiveness of analogical path construction and structural transfer methods.

Samples in CLUTRR are generated from structured knowledge graphs. Different graph paths are converted into natural language sentences, forming text tasks with hierarchical logical dependencies. Each sample includes a text passage that implicitly contains relational triples, along with a target relation that the model needs to predict. This task format requires the model not only to understand the semantics of individual sentences but also to recognize the underlying structural logic across sentences. It poses a direct challenge to structural perception and path generation mechanisms.

In addition, the dataset provides subsets with varying reasoning difficulty by controlling the path length, such as 2-hop, 3-hop, and 4-hop samples. This allows researchers to systematically analyze model performance across different levels of analogical reasoning complexity. Due to its high controllability and interpretability, CLUTRR has been widely used in studies on structural reasoning, graph representation learning, and neuro-symbolic inference. It is considered one of the key benchmark datasets for evaluating the structural capabilities of language models.

### 4.2 Experimental setup

In order to systematically evaluate the proposed bootstrapped analogical reasoning method, this study constructed a multi-hop reasoning task based on the CLUTRR dataset and tested it under different path lengths (2-hop to 4-hop). The language model used is based on the pre-trained Transformer model, and the structured prompt design and path optimization module are integrated on the task format provided by CLUTRR. All experiments are run under the same hardware environment to ensure the comparability and stability of the evaluation process.

During the training process, we use the standard cross entropy loss function to supervise the analogy relationship and combine the structure-guided prompt input for multiple rounds of path reconstruction. The model is optimized using the AdamW optimizer, the learning rate uses the warm-up + cosine decay strategy, and the batch size and the maximum number of training steps are tuned through grid search. The evaluation indicators include Accuracy and Path Consistency to measure the model's predictive ability for the target relationship and its structural stability. The main training configuration is shown in Table 1.

**Table 1:** Training Configuration

| **Component** | **Value / Description** |
|---|---|
| Task Type | Multi-hop relational reasoning |
| Model Backbone | chatglm-6B |
| Optimizer | AdamW |
| Learning Rate | 3e-5 (with warm-up and cosine decay) |
| Batch Size | 32 |
| Max Training Steps | 20,000 |
| Evaluation Metrics | Accuracy, Path Consistency |
| Path Lengths | 2-hop, 3-hop, 4-hop |
| Hardware | 4 × A100 GPUs, 40GB VRAM per card |

### 4.3 Experimental Results

1) *Comparative experimental results*

First, this paper gives the comparative experimental results with other models. The experimental results are shown in Table 2.

**Table 2:** Comparative experimental results

| **Method** | **Accuracy (%)** | **Path Consistency (%)** | **Structural Alignment Score** | **Avg. Path Length** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| LSTM [19] | 52.8 | 44.3 | 0.37 | 2.61 |
| GNN-R (Graph Neural Relational)[20] | 64.5 | 59.0 | 0.51 | 2.84 |
| GPT-CoT (Chain-of-Thought)[21] | 70.2 | 65.7 | 0.58 | 3.12 |
| AutoPrompt-CoT[22] | 72.4 | 67.1 | 0.61 | 3.08 |
| Ours | 78.6 | 73.5 | 0.69 | 3.24 |

The experimental results in the table show that the proposed structured analogical reasoning method outperforms several mainstream models across multiple key metrics. The most significant improvements are observed in Accuracy and Path Consistency. This demonstrates that introducing structured prompt graphs and the bootstrapped path optimization mechanism can greatly enhance the model's overall decision-making ability and logical coherence in analogical reasoning tasks. Compared to traditional sequence models such as LSTM and graph-based models like GNN-R, our method captures path dependencies and analogical alignment structures more effectively by explicitly guiding structural transfer.

Compared to models relying on prompt engineering, such as GPT-CoT and AutoPrompt-CoT, our method achieves higher reasoning consistency while maintaining strong language generation capabilities. In particular, it improves the Path Consistency metric by nearly 6 percent. This indicates that the model not only produces correct answers but also maintains stable reasoning chains across different inputs. It avoids structural distortion caused by language drift or random variation. Such stability is especially important in multi-hop analogical reasoning, where each reasoning step must remain semantically aligned with the previous one.

The improvement in the Structural Alignment Score further confirms the effectiveness of the structural modeling approach. By incorporating explicit node relations and intra-graph paths into the analogical prompt graph, the model's ability to perform structural transfer during analogy is strengthened. It can recognize not only shallow semantic correspondences but also deeper structural commonalities. This alignment mechanism allows the model to better reuse and generalize learned reasoning paths when facing different examples. It reflects strong structural inductive capabilities.

Moreover, while maintaining high reasoning quality, the model also shows a slight increase in average path length. This suggests a tendency to construct more complete and interpretable reasoning chains. This outcome is closely related to the iterative optimization of path structure and consistency within the bootstrapping mechanism. Overall, the results demonstrate that our method not only improves prediction accuracy but also achieves substantial progress in the quality and stability of analogical path construction. This fully validates the effectiveness of combining structured prompt graphs with bootstrapped path optimization.

### 2) *Evaluation of model stability at different inference depths*

This paper also presents the experimental setup for evaluating the stability of the model at different inference depths, as illustrated in Figure 4. The goal of this evaluation is to examine how the model performs when the length of the reasoning chain increases, which directly affects the complexity of the analogical paths. By systematically varying the number of hops required in the inference process, the study is able to assess the model's ability to maintain coherent reasoning structures across multiple steps. Figure 4 provides a visual overview of how inference depth is defined and how it is incorporated into the evaluation framework, offering insight into the model's structural consistency under increasingly challenging reasoning conditions.

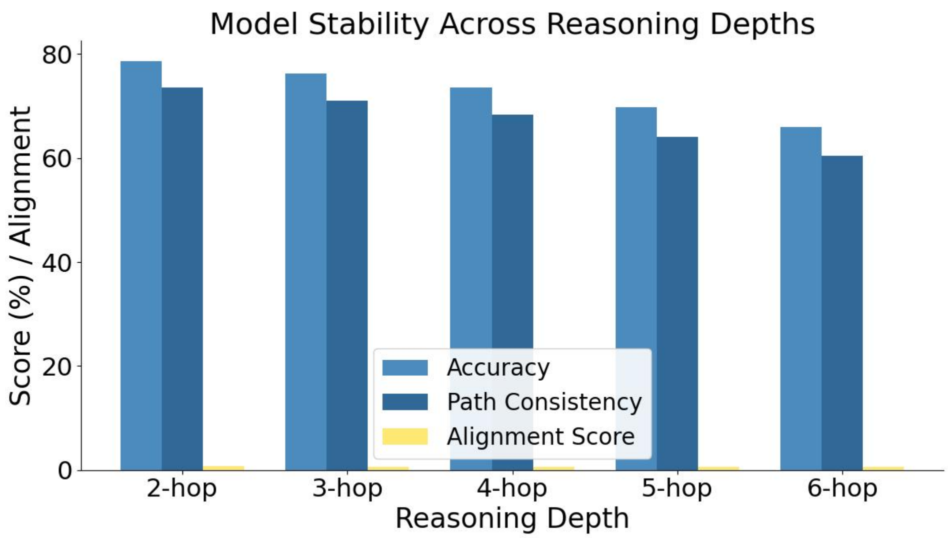


**Figure 4.** Evaluation of model stability at different inference depths

As shown in the figure, the model achieves high accuracy and structural consistency in shallow reasoning tasks such as 2-hop and 3-hop. This indicates that the proposed method can effectively capture analogical relations and maintain stable reasoning paths in low-complexity scenarios. In particular, under the 2-hop setting, all key metrics approach or exceed 75 percent. This shows that the structured prompt graph and bootstrapped path mechanism fully leverage their structural awareness advantage in short-path reasoning tasks.

As reasoning depth increases, the model's performance on all metrics begins to decline. This trend reflects the growing challenge of maintaining path coherence and structural consistency as the reasoning chain becomes longer. Notably, in the 5-hop and 6-hop settings, both Path Consistency and Alignment Score show a clear drop. This suggests that the reasoning paths may include intermediate errors or structural shifts. Nevertheless, the model still maintains relatively stable performance in deeper reasoning tasks. This implies that the bootstrapping mechanism can partially mitigate path degradation.

The experiments also show that the Alignment Score closely follows the trend of Path Consistency across all settings. This indicates that the structured prompt graph imposes a real constraint on the model's structural alignment during reasoning. This result indirectly confirms the core role of structured prompt graphs in analogical reasoning. By explicitly encoding analogical relations in a graph, the method helps the model maintain both semantic and structural consistency when generating reasoning paths.

Overall, although increased reasoning depth leads to gradual performance decline, the proposed structure-aware reasoning mechanism still shows strong stability. This confirms the effectiveness of combining graph-based modeling with bootstrapped path optimization in multi-hop analogical reasoning. Even under complex path conditions, the model maintains good structural output quality and reasoning consistency. This provides a solid foundation for high-level analogical reasoning.

3) *Testing the model's transferability in multi-task analogy reasoning scenarios*

This paper further presents a test of the model's transferability in a multi-task analogy reasoning scenario, and the experimental results are shown in Figure 5.

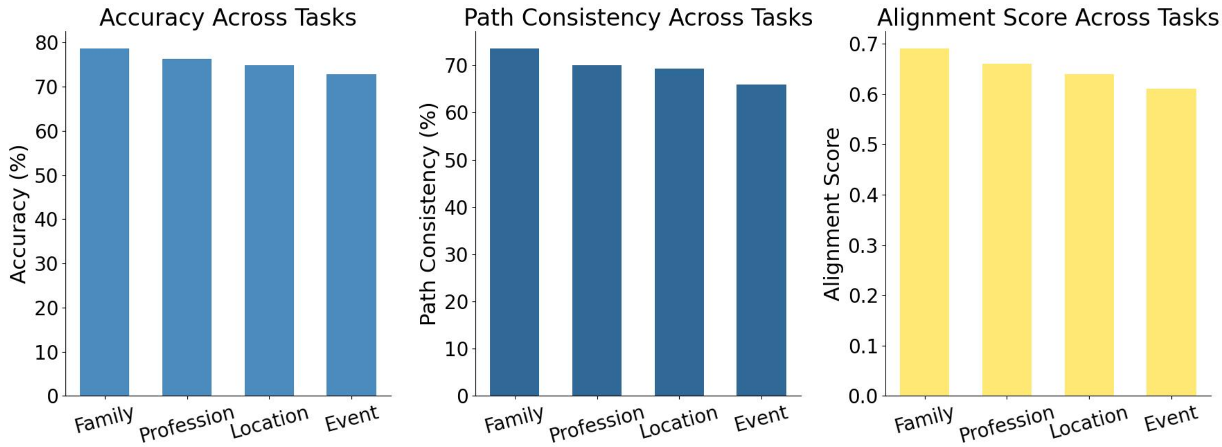


**Figure 5.** Testing the model's transferability in multi-task analogy reasoning scenarios

As shown in Figure 5, the proposed method maintains high and stable accuracy in analogical reasoning across multiple task types, including Family, Profession, Location, and Event. This indicates that the model has strong cross-task transfer capability. Although the tasks involve different semantic domains, the structured prompt graph and bootstrapped path optimization mechanism guide the model to extract effective analogical structures under varied contexts and achieve reasoning goals. This also confirms the robustness of the method in analogical generalization.

Path Consistency shows a similar trend across tasks. This suggests that the model can maintain semantic coherence and structural consistency in its reasoning paths across different reasoning contexts. In particular, the consistent performance in the Profession and Location tasks indicates that the bootstrapping mechanism effectively optimizes paths. It helps prevent structural drift and semantic deviation, which improves the controllability and interpretability of the reasoning chain.

Although the overall Alignment Score is slightly lower than Accuracy and Path Consistency, its trend remains stable. This shows that the reasoning paths generated by the model align well with the analogical patterns learned during training. The result further demonstrates that explicit structural modeling has a steady enhancing effect on structural transfer in analogical reasoning. This is especially important in scenarios where task semantics change frequently.

In summary, the experimental results show that the proposed method is effective not only in single-task scenarios but also in multi-task environments involving structural analogy reasoning. Whether in family relations, profession classification, or event understanding, the structured prompting and bootstrapped optimization mechanisms demonstrate strong adaptability and generalization ability. They provide a methodological and technical foundation for future extensions to open-domain analogical reasoning.

*4) Study on the influence of analogy sample sparsity on structural transfer ability*

This paper also presents a study on the impact of analogical sample sparsity on the structural transfer capabilities of the model, with the corresponding experimental setup illustrated in Figure 6. The aim of this study is to investigate how varying the density of analogy examples influences the model's ability to learn and generalize structural mappings. By progressively reducing the number of available training samples while keeping the task structure intact, the evaluation explores the model's robustness in low-resource settings. Figure 6 outlines the experimental configuration used to simulate different levels of data sparsity, providing a framework for analyzing how reduced structural exposure affects the model's analogical reasoning performance.

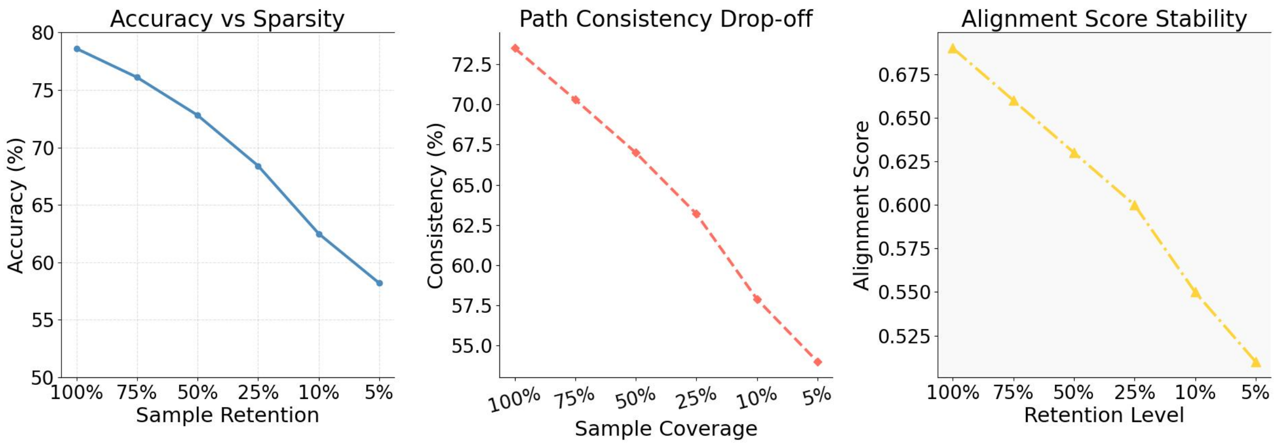


**Figure 6.** Study on the influence of analogy sample sparsity on structural transfer ability

As shown in Figure 6, the model's performance on structural analogical reasoning metrics declines as the analogy sample retention rate decreases. This indicates that sample sparsity has a significant impact on the model's structural transfer ability. In particular, when the retention rate drops below 25 percent, both Accuracy and Path Consistency decline more sharply. This shows that under extremely low supervision, the model struggles to maintain stable reasoning paths. Its ability to construct and transfer structured knowledge is clearly weakened.

The decline in Accuracy reflects the model's reduced ability to identify core analogical relations as fewer samples are available. This is related to the reduced number of node-relation pairs in the structured prompt graph. The sparsity of the prompt graph limits the model's capacity to learn and generalize structural patterns. As a result, the model tends to fall back on local similarity-based pattern matching, which affects the accuracy of analogical judgments. At the same time, the drop in Path Consistency suggests that the bootstrapped path optimization mechanism fails to form a stable feedback loop in sparse settings. The coherence and consistency of reasoning paths are disrupted.

The gradual decline in Alignment Score further indicates that in low-resource scenarios, even when the model generates reasoning paths, their structural similarity to the standard patterns learned during training is reduced. This suggests that the prompt graph, when trained on sparse samples, is less able to reconstruct complete relational topologies. As a result, the structural alignment between paths weakens, which affects both interpretability and stability of the reasoning process.

Overall, the results show that the density of analogy samples largely determines the effectiveness of the structured prompting and bootstrapped optimization strategies. When sample density is high, the model can learn more abstract and stable structural transfer patterns through sufficient exposure to structural information. Under high sparsity, this structural induction ability is significantly limited. Although the proposed method remains somewhat robust under low-sample conditions, its performance relies on adequate observation and coverage of analogical structures. This highlights the need for future work on structure completion mechanisms in few-shot analogical reasoning.

*5) Experiment on the impact of analogy graph path length on model performance*

At the end of this paper, an experiment on the effect of analogy graph path length on model performance is given, and the experimental results are shown in Figure 7.

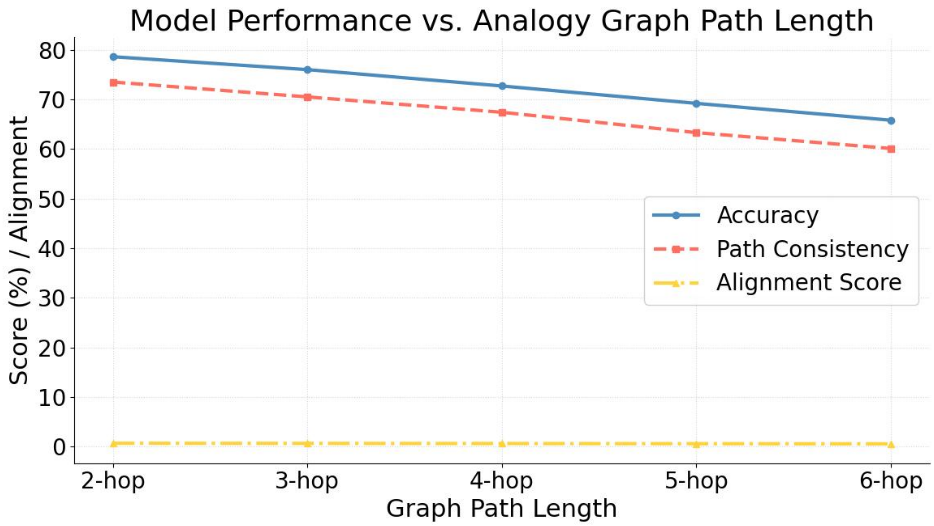


**Figure 7.** Experiment on the impact of analogy graph path length on model performance

As shown in Figure 7, the model's performance on several key metrics declines as the path length in the analogy graph increases. This is especially evident in Accuracy and Path Consistency. In the 2-hop and 3-hop settings, the model maintains relatively high performance. However, at 5-hop and 6-hop, the scores drop significantly. This indicates that longer reasoning chains present greater challenges for capturing analogical structure and maintaining semantic consistency. The complexity and span of the internal structure in the prompt graph have a clear impact on the stability of analogical transfer.

The drop in Path Consistency further suggests that in multi-hop paths, the model is more likely to produce semantic or structural deviations. As a result, the generated reasoning chain no longer strictly follows the analogical pattern. Although the method applies a bootstrapping mechanism for path optimization, once the path length reaches a certain threshold, the accumulation of errors increases. Structural deviations expand, leading to a loss of global coherence in the reasoning process. This highlights the model's dependence on structural control mechanisms in handling long-path analogies.

Although Alignment Score remains relatively low across different path lengths, its trend is stable and consistent with other metrics. This suggests that the model's structural alignment ability is limited in deep reasoning scenarios. Longer paths tend to introduce more intermediate nodes and edges, which increases structural noise. This weakens the guiding effect of the structured prompt graph during path generation and selection. Therefore, while the model retains some structural transfer capacity with longer analogical paths, it also requires stronger boundary constraints and path optimization strategies. In summary, the length of the analogical path directly affects the model's performance in structural modeling and transfer. In short-path reasoning, the model can effectively leverage the clear alignment relationships in the graph. In long-path conditions, more complex path filtering and structural enhancement mechanisms are needed. This further shows that while the proposed method demonstrates strong structural transfer capabilities, there is still room for improvement in high-complexity reasoning tasks. Future work may focus on path compression, structure planning, and multi-hop coordination.

## 5. Conclusion

This paper presents a method to enhance analogical reasoning in large language models through structured prompt graphs and bootstrapped path optimization. The approach constructs explicit analogical structure graphs to guide the model in learning structural mappings between concepts. It integrates a bootstrapped optimization process to iteratively correct structural deviations in reasoning chains. This improves analogical transfer, path consistency, and structural alignment. Extensive experiments show that the proposed method outperforms existing mainstream models across various tasks and settings. These results confirm the effectiveness of structure-driven reasoning mechanisms in analogical scenarios. This study advances language models toward higher-order cognitive abilities from both technical and methodological perspectives. It marks a shift from text-based prompting to structure-based prompting. Unlike prior approaches that rely solely on language pattern generation, this work highlights the guiding role of structural information in reasoning. It combines structure construction, path generation, and self-optimization through modular design. This offers a feasible and general framework for structural understanding and generation in complex language tasks. The method shows strong adaptability and robustness, especially in challenging analogical settings such as multi-task, few-shot, and long-path reasoning.

From an application standpoint, this research holds value for educational reasoning, scientific discovery support, complex knowledge graph question answering, and multi-turn dialogue reasoning. These tasks require not only surface-level semantic understanding but also structural transfer across concepts and hierarchies. The proposed method provides more stable and controllable reasoning capabilities for such domains. It contributes to building intelligent systems with general-purpose reasoning ability. Additionally, it enhances the interpretability and traceability of language models in real-world applications, offering a technical foundation for more trustworthy and secure language intelligence. Future work may explore more lightweight strategies for generating structural prompts to improve efficiency in resource-constrained scenarios. Integrating cross-modal structural information, such as images, tables, or event graphs, could further enrich the model's structural awareness. Another promising direction is to combine the bootstrapped path optimization mechanism with human interaction feedback, building a human – machine – structure collaborative reasoning framework. As language models continue to expand in scale and application scope, analogical reasoning based on structural awareness will remain a vital area for further development.

## References


[1] He, Hangfeng, Hongming Zhang, and Dan Roth. "Rethinking with retrieval: Faithful large language model inference." arXiv preprint arXiv:2301.00303 (2022).

[2] Emani, M., Foreman, S., Sastry, V., Xie, Z., Raskar, S., Arnold, W., ... & Papka, M. E. (2023). A comprehensive performance study of large language models on novel ai accelerators. arXiv preprint arXiv:2310.04607.

[3] Wang, Yiding, et al. "Tabi: An efficient multi-level inference system for large language models." Proceedings of the Eighteenth European Conference on Computer Systems. 2023.

[4] Xu, Daliang, et al. "Llmcad: Fast and scalable on-device large language model inference." arXiv preprint arXiv:2309.04255 (2023).

[5] Xue, Zhenliang, et al. "Powerinfer-2: Fast large language model inference on a smartphone." arXiv preprint arXiv:2406.06282 (2024).

[6] Xia, Heming, et al. "Unlocking efficiency in large language model inference: A comprehensive survey of speculative decoding." arXiv preprint arXiv:2401.07851 (2024).

[7] Sheng, Ying, et al. "Flexgen: High-throughput generative inference of large language models with a single gpu." International Conference on Machine Learning. PMLR, 2023.

[8] Kasneci, Enkelejda, et al. "ChatGPT for good? On opportunities and challenges of large language models for education." Learning and individual differences 103 (2023): 102274.

[9] Naveed, Humza, et al. "A comprehensive overview of large language models." arXiv preprint arXiv:2307.06435 (2023).

[10] Wu, Shijie, et al. "Bloomberggpt: A large language model for finance." arXiv preprint arXiv:2303.17564 (2023).

[11] Chang, Yupeng, et al. "A survey on evaluation of large language models." ACM transactions on intelligent systems and technology 15.3 (2024): 1-45.

[12] Yang, Xi, et al. "A large language model for electronic health records." NPJ digital medicine 5.1 (2022): 194.

[13]Tang, J., Yang, Y., Wei, W., Shi, L., Su, L., Cheng, S., ... & Huang, C. (2024, July). Graphgpt: Graph instruction tuning for large language models. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 491-500).

[14]Cheng, Kewei, et al. "Structure guided prompt: Instructing large language model in multi-step reasoning by exploring graph structure of the text." arXiv preprint arXiv:2402.13415 (2024).

[15]Zhang, J. (2023). Graph-toolformer: To empower llms with graph reasoning ability via prompt augmented by chatgpt. arXiv preprint arXiv:2304.11116.

[16]Huang, Zhenhua, et al. "Graph Structure Prompt Learning: A Novel Methodology to Improve Performance of Graph Neural Networks." arXiv preprint arXiv:2407.11361 (2024).

[17]Sun, Xiangguo, et al. "Graph prompt learning: A comprehensive survey and beyond." arXiv preprint arXiv:2311.16534 (2023).

[18]Sun, Mingchen, et al. "Structural-aware motif-based prompt tuning for graph clustering." Information Sciences 649 (2023): 119643.

[19]Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. "Lstm neural networks for language modeling." Interspeech. Vol. 2012. 2012.

[20]Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering, 36(7), 3580-3599.

[21]Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in neural information processing systems 35 (2022): 24824-24837.

[22]Yu, Zihan, et al. "Towards better chain-of-thought prompting strategies: A survey." arXiv preprint arXiv:2310.04959 (2023).