

# Metapath of Thoughts: Verbalized Metapaths in Heterogeneous Graph as Contextual Augmentation to LLM

Harshvardhan Solanki  
hvsai@amazon.com  
Amazon Web Services  
Seattle, WA, USA

Yihui Chong  
cyihui@dso.org.sg  
DSO National Laboratories  
Singapore

Jyoti Singh  
singjyo@uw.edu  
University of Washington, Tacoma  
Tacoma, WA, USA

Ankur Teredesai  
ankurt@uw.edu  
University of Washington / CueZen Inc.  
Seattle, Washington, USA

**Keywords:** Text-Rich Network, Graph Transformer, Large Language Models, Explainable AI

## 1. Abstract

Heterogeneous graph neural networks (HGNNs) excel in capturing graph topology and structural information. However, they are ineffective in processing the textual components present in nodes and edges and thus producing suboptimal performance in downstream tasks such as node-classification. Additionally, HGNNs lack in their explanatory power and are considered black-box. Although, Large Language models (LLMs) are good at processing textual information, however, utilizing them for tasks like node-prediction can be non-trivial since it is difficult to identify the ideal graphical context and present it in a form suitable for LLMs to consume effectively. We introduce a framework that combines the strengths of both models by leveraging the context obtained through metapaths, which are generated during the training of HGNNs. This approach enables the understanding of complex and indirect relationships between different types of nodes. Our novel framework enhances the prediction accuracy of HGNNs and the transparency of their decision-making process through natural language explanations provided by LLMs. We demonstrate that our proposed framework outperforms FASTGTN (SOTA on heterogeneous node classification tasks), an HGNN tailored for heterogeneous graph data, on two network datasets (DBLP citation graph and Goodreads graph dataset), with improvements in F1 score from 0.81 and 0.66 of the baseline to 0.9 and 0.91, respectively. Furthermore, the efficacy of the framework in generating explanations has been evaluated through human evaluation, considering metrics such as helpfulness and factual correctness.

## 2. Introduction

Heterogeneous graphs (HGs) are typically characterized by their diversity in types of nodes and edges. They are used

to represent complex relationships and entities in various domains such as natural language processing [44] [34], community detection [29] [6], recommendation systems [20] [33] and biological networks. This heterogeneity enables a nuanced and comprehensive representation of real-world data, significantly enhancing the capability of graph analysis in capturing intricate relationships and interactions [40]. Graph neural networks (GNNs) are widely used as a tool in representation learning on graph-structured data, demonstrating efficacy in tasks such as node classification and link prediction. [39] However, a notable limitation of most Graph Neural Networks (GNNs) is their reliance on a fixed and homogeneous graph structure for the operations in their training algorithm. A proposed solution involves the manual construction of meta-paths, which consist of sequences linked by heterogeneous edges, to convert a heterogeneous graph into a homogeneous graph as defined by these meta-paths. Since the manual construction of metapaths requires domain experts and the accuracy of any downstream task is highly dependent on the choice of these metapaths, thus, Heterogeneous Graph Neural Networks (HGNNs) such as GTN (Graph Transformer Networks) and FastGTN have been developed to autonomously transform a heterogeneous input graph into functional meta-path graphs tailored for specific tasks, while simultaneously learning node representations on these graphs in an integrated, end-to-end manner for predictive tasks[39]. Heterogeneous Graph Neural Networks (HGNNs) [43] are widely used for tasks like node classification, link prediction [9] [32], and recommendation generation [20] [33], by effectively aggregating and learning from the information contained within HGs with the help of meta paths [11], that enable the extraction and understanding of complex and indirect relationships between different types of nodes [7] [32]. However, due to the complex and non-linear transformations in HGNNs, these model often lack in their explainability power and are considered a black box [38]. Understanding the reasoning behind their predictions or decisions remains an open area of research, posing challenges

to transparency and trust, especially in critical applications [2] [15].

Large Language Models (LLMs) have recently made significant strides in tasks involving reasoning and utilizing general knowledge, largely due to their ability to process and generate human-like text [19]. These models understand context, infer meaning from complex sentences, and generate coherent and contextually relevant text [21]. This makes them particularly effective in tasks requiring reasoning [5], such as answering questions [28], summarizing texts [41], or creating content [18] [1]. Even with the exemplary reasoning abilities of LLMs in various domains, their usage with structured data, particularly in the form of heterogeneous graphs, remains largely under-explored. While Heterogeneous Graph Neural Networks (GNNs) effectively capture and utilize the structural and relational information inherent in graphs (like social networks, citation networks, etc.), LLMs primarily deal with unstructured textual data. Combining these two to leverage both textual understanding and structural learning, especially for tasks related to heterogeneous graphs like consuming meta-path-aware context for predictions and explanation, is an area that is still an active area of research. The primary challenge lies in aligning the fundamentally different data modalities (textual vs. graph-structured) and ensuring that the combined model can generate predictions and explanations that are both contextually relevant and aware of the intricate relationships in heterogeneous graph data, which is a proven strength of HGNNs. This issue highlights the need for a framework that can facilitate reasoning by providing a metapath-based context tailored to the unique structure of HGNNs so as to utilize the combined power of LLMs and HGNN.

Previous works have extensively explored Graph Neural Networks (GNNs) and their explainability. Studies by [16] [37] have provided valuable insights into the effectiveness of different types of explainers. On a broad level, these explainers are categorized as gradient-based [26] [27] [22], perturbation-based [36] [17] [8], decomposition-based [3] [23], and surrogate-based models [10] [42] [30]. Particularly, the role of meta-paths and node or edge masks in elucidating the decision-making process of GNNs has been a central theme. However, despite these advancements, the field still faces significant challenges, particularly in the context of heterogeneous graphs (HGs). The intricacies of HGs, with their diverse node and edge types, demand more nuanced explanation methods that can handle the complexity and varied nature of the data.

Our framework focuses on improving the predictions and providing explainability, emphasizing the topological importance within heterogeneous graph-structured data. Central

to our approach are metapaths, which are critical in the classification tasks of HGNNs. These learned metapaths are used to craft explanations and enhance the interpretability of these networks. The key lies in how the context is constructed, utilizing learned metapath-based instances along with semantically similar and randomly picked instances to enrich the context. Our classification scores surpass those of the FastGTN baseline, a specialized network for heterogeneous graph-structured data. Furthermore, the generated explanations are preferred by human labelers for their helpfulness and truthfulness over those produced solely on homogeneous context—i.e., explanations that neglect graph heterogeneity and rely only on n-hop neighbors.

We also address the challenge of manually creating metapath-specific prompts for any dataset and propose a novel Dynamic Prompt Generation framework comprising of verbalized metapaths as thoughts.

The key contributions of this work are:

1. **Enhanced Node Classification Method:** Our approach significantly improves node classification performance, outperforming the baseline FastGTN model (SOTA on heterogeneous node classification tasks)[39] when considering f1 score as the metric.
2. **Dynamic Prompt Generation Framework:** To overcome the challenge of manually creating metapath-specific prompts for each dataset, we introduce a novel prompt generation framework that automatically extracts metapath types and verbalize them in natural language to generate context-specific prompts for our LLMs. The framework generates an automated prompt with context enriched not just by metapath instances but also by instances obtained through semantic similarity to make the prompts contextually richer.
3. **Development of a Sophisticated Explanation Framework:** Focused on heterogeneous graph-structured data, this framework is adept at providing contextually relevant, and easily interpretable explanations. Through these contributions, our work stands out in its ability to blend advanced data processing techniques with user-centered design, making complex graph-structured data more accessible and understandable.

### 3. Background

#### Definition 1: Heterogeneous Graph [24] [39]

A directed graph  $G = (V, E, T_v, T_e)$  is defined as a heterogeneous graph if each node  $v \in V$  and each edge  $e \in E$  have corresponding type mapping functions  $\tau_v(v) : V \rightarrow T_v$  and  $\tau_e(e) : E \rightarrow T_e$ . This type of graph can be described using a set of adjacency matrices  $\{A_t\}_{t=1}^{|T_e|}$

or a tensor (i.e.,  $A \in \mathbb{R}^{|V| \times |V| \times |T_e|}$ ), where  $A_t \in \mathbb{R}^{N \times N}$  represents the adjacency matrix for the  $t$ -th edge type, and  $|V| = N$ . In this context,  $A_t[i, j]$  indicates the weight of the edge of type  $t$  from node  $j$  to node  $i$ . When there is only one type of node and one type of edge, i.e.,  $|T_v| = 1$  and  $|T_e| = 1$ , the graph is referred to as a homogeneous graph.

**Definition 2: Meta-paths [32] [39]**

In the context of heterogeneous graphs, a meta-path refers to a sequence of edges that connect nodes through multiple types of relations. Formally, a meta-path can be represented as:

$$\mathcal{P} = v_1 \xrightarrow{t_1} v_2 \xrightarrow{t_2} \dots \xrightarrow{t_{n-1}} v_n,$$

where  $v_1, v_2, \dots, v_n$  denote the nodes in the sequence, and  $\xrightarrow{t_i}$  represents an edge of type  $t_i$  from node  $v_i$  to node  $v_{i+1}$ . Here,  $t_i$  specifies the edge type of the edge  $(v_i, v_{i+1})$  along the meta-path.

A meta-path can also be described as a composite relation  $R = t_1 \circ t_2 \circ \dots \circ t_n$  between node  $v_1$  and  $v_{n+1}$ , where  $R_1 \circ R_2$  denotes the composition of relations  $R_1$  and  $R_2$ . Given the composite relation  $R$  or the sequence of edge types  $(t_1, t_2, \dots, t_n)$ , the adjacency matrix  $A_P$  of the meta-path  $P$  is derived from the multiplication of adjacency matrices as  $A_P = A_{t_n} \dots A_{t_2} A_{t_1}$ .

The concept of a meta-path includes multi-hop connections and introduces new graph structures represented by adjacency matrices. For instance, the meta-path Author-Paper-Conference (APC),

$$A \xrightarrow{AP} P \xrightarrow{PC} C$$

which can be represented as  $A \xrightarrow{AP} P \xrightarrow{PC} C$ , results in an adjacency matrix  $A_{APC}$  by multiplying  $A_{AP}$  and  $A_{PC}$ .

**Definition 3: Metapath Instance [12]**

An instance  $p$  of the meta-path  $P$  is a particular sequence of nodes  $\langle n_1, n_2, \dots, n_l \rangle$  within the network that adheres to the pattern specified by  $P$ . Multiple instances can correspond to a single metapath.

For example, in a citation network, consider the meta-path "Author-Paper-Author" (APA). Various sequences of nodes, such as  $A_i P_i A_j$ , can fit this metapath pattern, making them instances of the "APA" metapath.

**Heterogenous Graph Neural Network and FastGTN[40]**

Heterogeneous graph neural networks (HGNNs) are advanced models designed to handle the complexity of heterogeneous graphs (HGs), which consist of diverse types of nodes and edges. Unlike traditional graph neural networks, HGNNs account for the semantic and structural heterogeneity inherent in HGs. They use deep learning

techniques to learn embeddings from node attributes and interactions, effectively capturing the rich, multi-faceted information within these graphs. This capability makes HGNNs particularly powerful for tasks requiring a nuanced understanding of complex networks. The training methodology [39] employed by Graph Transformer Networks (GTN) and FastGTN involves the transformation of original graphs into multiple subgraphs referred to as metapaths above. This transformation is integral to the learning process, where node representations are further derived through convolution operations conducted on top of these metapaths. Learnt node representations on top of these metapaths are further used for downstream tasks like node classification. Additionally, the methodology used in FastGTN also provides the weighting/attention-score on how it influences the node prediction. They also demonstrated that meta-paths predefined by domain knowledge consistently receive top rankings from FastGTNs, further validating their influence through the use of attention scores.

**Explainability of Supervised Machine Learning [4]**

Consider a supervised task  $T$  with the aim of learning a mapping from  $X$  to  $Y$ , and a model  $M$  trained for this task. Given a set of  $(x, y)$  pairs  $\subseteq (X, Y)$  and the model  $M$ , generate an explanation  $e$  from a given set  $D_E$  such that  $e$  "explains" the prediction  $\hat{y} = M(x)$ . These explanations can be either local to a single test input  $(x, y)$  or global when they explain prediction over a specific dataset  $D' \subseteq (X, Y)$ . Further, the explanation can be generated either post-hoc (i.e., after the model training) or ante-hoc where the model itself is self-interpretable, i.e., it explains its predictions.

Local explanations only consider the local neighborhood of the given data instance while global explanations are concerned about the model's overall behavior and thus, search for patterns in the model's predictions. In this work, we focus on Local explanations.

**GNNs Explainability**

The survey [14] presents a comprehensive overview of GNN explainability techniques. There exists a research gap for techniques that generate natural language explanations that are crucial to the practical application of GNN utilizing metapaths, which are crucial for interpreting heterogeneous graph neural networks (HGNNs). Metapaths [11] represent sequences of node types and edge types in heterogeneous graphs, offering a pathway to understand complex relationships and interactions within the data.

Most current explainability methods discussed in [14] focus on visual or quantitative interpretations, such as highlighting important nodes and edges or providing importance scores. These methods, while useful, fall short of offering intuitive, human-readable explanations that describe the reasoning behind a model’s prediction in natural language. The integration of metapaths into generative natural language explanations would bridge this gap by translating the intricate relationships within heterogeneous graphs into comprehensible narratives, enhancing interpretability for non-expert users.

### Use of LLMs for model interpretability and natural language explanations

The work presented in [25] proposes the opportunity to rethink LLMs for interpretability with a more ambitious scope. LLMs can elicit more elaborate explanations than the previous generation of interpretable ML techniques. While previous methods have often relied on restricted interfaces such as saliency maps, LLMs can communicate directly in expressive natural language.

#### 4. Proposed methodology

The approach involves providing a combination of context coming from verbalized representation of meta-paths as well as context based on semantic similarity.

Figure 1, depicts the detailed overview of the proposed methodology. We explain these steps in detail in the following subsections:

- a. These meta-paths are created as part of the training of fastGTN for the purpose of node classification. See [39] for the training of fastGTN and meta-path computation.
- b. Top-K metapath types are chosen among the total metapath types based on the descending order of the weights on how these influence the node classification. Chosen metapath types are then further filtered based on whether they start with the target node type or not.
- c. For an LLM to comprehend and make use of metapaths, these need to be verbalized and summarized in natural language, this is done using a one-shot LLM call, LLM is prompted to generate a verbal interpretation of a given metapath type. This involves presenting the LLM with a structured prompt consisting of node-type, edge-type, and the meaning of different edges and asking it to produce a logical and meaningful explanation. For the verbalization of metapath, refer to figure 2 for verbalization instructions prompt template, figure 3 for the augmented prompt and 4 for example generation.

- d. Depending on the local graph topology of the target node, multiple instances of each metapath type are possible. To construct the contextual representation, instances of distinct metapath types are sampled based on the normalized weights assigned to the corresponding metapath types. The verbal interpretation derived from the initial analysis must then be applied to all instances of a given metapath type through the utilization of regular expressions.
- e. As the datasets and use cases pertaining to heterogeneous graphs often involve nodes with textual attributes, the textual content associated with each node participating in the metapath instances is additionally incorporated alongside the verbalized representation of the metapath. Refer to figure 7 for the format of the metapath context along with textual attributes
- f. Since the contextual information derived solely from the metapath may be insufficient, we also consider the textual content associated with nodes exhibiting high semantic similarity to the target node under consideration.
- g. Subsequent to providing the contextual information, a series of instructions is presented in a step-by-step, chain-of-thought manner to establish a connection between the contextual information derived from the graph and the target node. This process is designed to elicit predictions and explanations from the LLM by leveraging the provided context. Refer to figure 5 for the CoT style instructions for prediction and explanation.

In summary, our LLM prompting strategy involves a two-step process: first, generating verbal interpretations of metapaths, and second, using these interpretations to build a context for accurate prediction and explanation. This approach ensures that the LLMs can effectively leverage the metapaths for our classification and explanation tasks.

#### 5. Dataset description

We evaluate our model on 2 public heterogeneous graph datasets: We use one citation network dataset DBLP [13], and a books dataset Goodreads [31] [?]. Though the raw datasets consisted of multiple labels, for our analysis we restricted to just using three labels. The labels of target node type books in Goodreads are “fiction”, “non-fiction”, and “romance”, whereas in DBLP the labels in target node type of paper are “machine learning”, “computer networking”, and “theoretical computer science”. The statistics of the filtered datasets and the associated heterogeneous graphs used in our experiments are shown in Table 1.

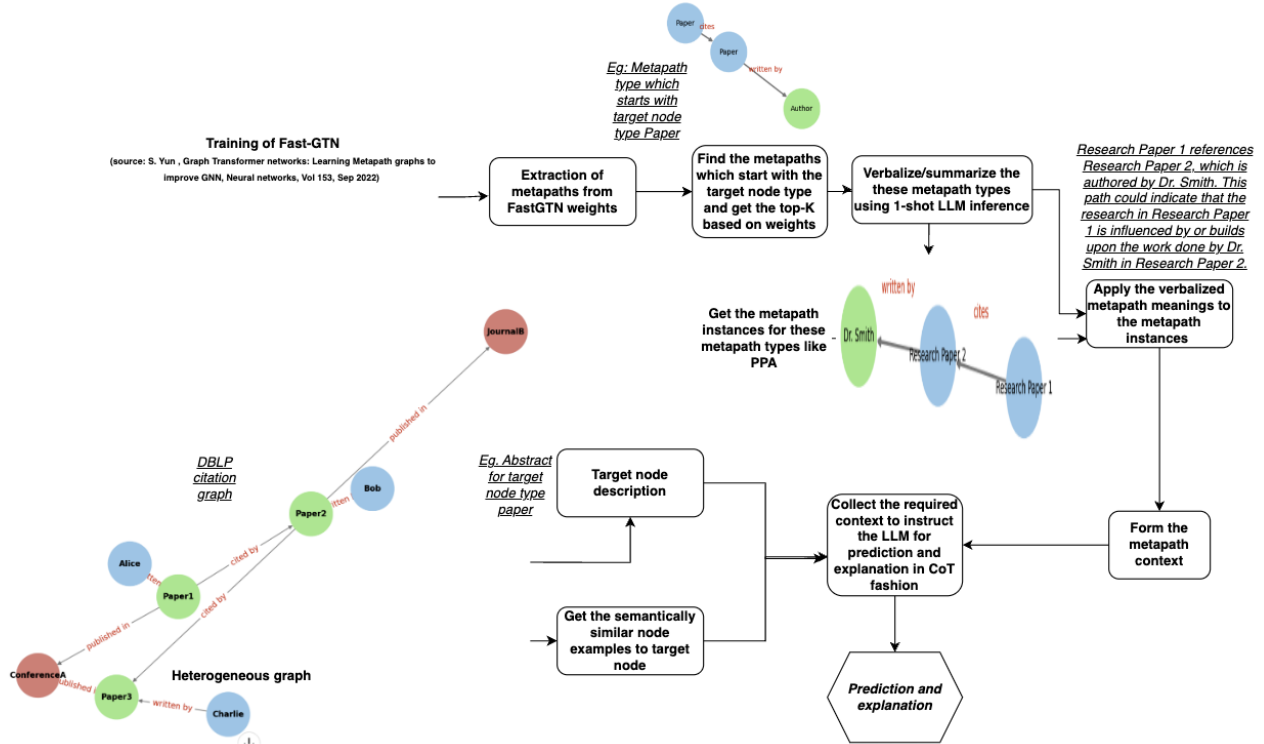


Figure 1. Proposed methodology

DBLP consists of three types of nodes (papers (P), authors (A), conferences (C)), four types of edges (PA, AP, PC, CP), and field of study for papers as labels. The Goodreads dataset contains five types of nodes (book (B), author (A), format (F), publisher (P), language code (L)), and theme of books as labels. Both datasets are characterized by textual attributes for specific node types: the Goodreads dataset includes book descriptions as the textual component for its 'book' node type, while the DBLP dataset includes abstracts as the textual component for its 'paper' node type. Other node types in both datasets do not possess textual components except their title.

DBLP dataset as used in the FastGTN [39] was not textually attributed in the same way as ours since their version of the DBLP dataset had only the title of the nodes as the textual component. We selected textually attributed graphs due to their widespread occurrence in real-world scenarios. [35]. The main statistics of the two datasets are summarized in Table 1.

## 6. Experimental setup

### 6.1 Comparison with other approaches

**Homogeneous Context** For checking whether the heterogeneous context of the graph as provided in the form

| Dataset   | Node Types         | Edge Types                |
|-----------|--------------------|---------------------------|
| DBLP      | Author: 17,567     | Author-Paper: 18,810      |
|           | Conference: 1,562  | Conference-Paper: 5,687   |
|           | Paper: 6,000       | Paper-Author: 18,810      |
|           |                    | Paper-Conference: 5,687   |
|           |                    | Paper-Paper: 261          |
| Goodreads | Author: 4,452      | Author-Book: 4,762        |
|           | Book: 3,600        | Book-Author: 4,762        |
|           | Publisher: 1,790   | Book-Language code: 2,040 |
|           | Format: 26         | Language code-Book: 2,040 |
|           | Language code: 47  | Book-Publisher: 2,665     |
|           |                    | Publisher-Book: 2,665     |
|           |                    | Book-Book: 32             |
|           | Book-Format: 2,762 |                           |
|           | Format-Book: 2,762 |                           |

Table 1. Dataset statistics for DBLP and Goodreads.

of a metapath is adding any value, we ignore the heterogeneity aspects of the graph structure and treat it as a homogeneous graph. This approach allowed us to assess the classification performance and the quality of the explanations obtained from the language model (LLM) without considering the heterogeneous context.

We tested different levels of neighborhood hops by taking random neighbors at 1, 2, and 3 hops for our use case. By extensively evaluating different hop levels, we aimed to validate the significance of heterogeneity and the potential advantages of a meta-path-based approach. For the approach based on 1 hop, we randomly selected  $N$  samples from the immediate neighborhood (one hop away). For the 2-hop approach, we randomly selected  $N/2$  neighbors from the first hop and  $N/2$  instances from the second hop. Similarly, for the 3-hop approach, we randomly selected  $N/3$  neighbors from the first, second, and third hop respectively.

The homogeneous context obtained from these different hop levels was then passed as input to our LLM prompt. By analyzing the classification performance and the quality of the explanations generated by the LLM using this homogeneous context, we aimed to assess the potential benefits of incorporating heterogeneous graph structures and the meta-path-based approach in our analysis.

**Using LLM with examples of labels as the only context** Since there can be leakage of information as LLMs are trained on the internet corpus, so we would like to check the performance of prediction and explanation by using LLM with examples as the only context. Additional gain on top of this approach will tell us the lift for using alternate mechanisms of context.

### Task 1: Node Classification task

In our experiments, we evaluated the performance of our node classification models on two distinct datasets DBLP and Goodreads, each with specific characteristics and challenges. Table 4 of the appendix gives details about LLM Inference Hyperparameters, Table 6 gives details about Fast-GTN Training Hyperparameters, Table 8 gives details about dataset Label Types and Distribution used in Experiment.

Considering the latency and the cost of tokens, we restricted the number of metapath instances to a fixed number (here 40), and for an equivalent comparison to the approach involving homogeneous context, the number of neighbors has been restricted to the same number. Table 7 gives details about Hyperparameters used in Homogeneous Context and Table 5 for the Meta-Path Hyperparameters.

### Task 2: Explanation Generation Ability

**Explanation Metrics: Human Evaluation:** The proposed methodology demonstrated superior performance compared to the FastGTN and other alternative approaches, such as utilizing LLMs with only examples as context or employing homogeneous context when considering the prediction task. This outcome indirectly suggests that an approach exhibiting better prediction capabilities should also provide more accurate and informative explanations. We would like to measure the quality of explanation in both absolute as well as relative terms. For a relative comparison, we would like to restrict ourselves to comparing only with the homogeneous context since its prediction performance is closer compared to the proposed methodology. Explanation was compared for these approaches when both of them gave the same prediction.

Human evaluators are shown explanations from approaches 1-3 and asked to rank them based on their preferences. Post-ranking, evaluators rate each explanation on a scale of 1-3, considering aspects such as helpfulness (with 3 being the highest and 1 the lowest).

Helpful Explanation Label Definitions for Human Labelling:

#### a. Poor

- Clarity: Extremely confusing with excessive jargon or complex language.
- Relevance: Fails to address the main question or topic, often veering off-topic.
- Depth: Lacks meaningful insight, oversimplifies, or omits crucial information.

#### b. Partial

- Clarity: Understandable but lacks clarity in parts.
- Relevance: Addresses the topic but includes irrelevant information or misses key aspects.
- Depth: Provides basic information but lacks important details and insights.

#### c. Complete

- Clarity: Very clear, with well-articulated concepts and points.
- Relevance: Highly relevant and focused, with little to no irrelevant information.
- Depth: Comprehensive understanding, including important details and nuanced insights.

| Dataset   | Approach  | Micro_f1       | Macro_f1       |
|-----------|---|----------------|----------------|
| DBLP      | Fast-GTN  | 0.81           | 0.82           |
|           | Only LLM with 5 shot examples per label   | 0.767          | 0.793          |
|           | LLM with context up to one-hop and 5 shot random examples per label   | 0.721          | 0.745          |
|           | LLM with context up to two-hop and 5 shot random examples per label   | 0.78           | 0.81           |
|           | LLM with context up to three-hop and 5 shot random examples per label   | 0.799          | 0.825          |
|           | LLM with metapath as context and 5 shot random examples per label   | 0.876          | 0.882          |
|           | <b>LLM with metapath as context and 2-shot semantic examples per label and 3-shot random examples per label</b> | <b>0.90259</b> | <b>0.90533</b> |
| Goodreads | Fast-GTN  | 0.667          | 0.672          |
|           | Only LLM with 5 shot examples per label   | 0.771          | 0.791          |
|           | LLM with context up to one-hop and 5 shot random examples per label   | 0.84           | 0.853          |
|           | LLM with context up to two-hop and 5 shot random examples per label   | 0.843          | 0.856          |
|           | LLM with context up to three-hop and 5 shot random examples per label   | 0.855          | 0.8606         |
|           | LLM with metapath as context and 5 shot random examples per label   | 0.877          | 0.886          |
|           | <b>LLM with metapath as context and 2-shot semantic examples per label and 3-shot random examples per label</b> | <b>0.904</b>   | <b>0.909</b>   |

Table 2. Comparison of different approaches on DBLP and Goodreads datasets.

| Dataset   | % of labels with factual correctness rating as 3 | % of labels with Helpfulness rating as 3 | % of labels with Helpfulness rating as 2 |
|---|--|--|--|
| <b>DBLP</b>   |  |  |  |
| LLM with homogeneous context  | 95.12  | 10                                       | 90                                       |
| LLM with metapath as well as semantic examples (proposed methodology) | 97.5   | 30                                       | 70                                       |
| <b>Goodreads</b>  |  |  |  |
| LLM with homogeneous context  | 94.23  | 19.23                                    | 76.92                                    |
| LLM with metapath as well as semantic examples (proposed methodology) | 96.15  | 38.46                                    | 57.69                                    |

Table 3. Comparison of Factual Correctness and Helpfulness Ratings between Different Approaches for DBLP and Goodreads Datasets

Factual correctness explanation Label Definitions for Human Labelling:

a. **Incorrect**

- Accuracy: Contains factual errors or unsubstantiated claims.
- Consistency: Contains contradictions or statements that are inconsistent with verified data.

b. **Partially Correct**

- Accuracy: Mostly accurate but includes minor errors or oversimplifications.
- Consistency: Generally consistent but has occasional discrepancies that need clarification.

c. **Fully Correct**

- Accuracy: No factual errors, fully accurate throughout.
- Consistency: Completely consistent with current, verified data and information.

These metrics and methods enable us to rigorously evaluate the explanation generation capabilities of different approaches, ensuring a blend of objectivity and human-centric assessment. In order to perform the human evaluation for both of the datasets i.e. Goodreads and DBLP, we had setup AWS Sagemaker ground truth jobs and hired Human labelers, who performed the ranking exercise independently.

## 7. Results

### Task 1: Node Classification

Tables 2 demonstrate the results of different methods used for node classification. In this study, we evaluated the performance of our proposed method on two citation network datasets, DBLP and Goodreads. Our results demonstrate a significant improvement in the micro F1 score compared to the baseline FastGTN method (SOTA on heterogeneous node classification task). Specifically, on the DBLP dataset, our method achieved a micro F1 score of 0.9, representing an increase from 0.81 obtained by FastGTN. Similarly, on the Goodreads dataset, our method exhibited a micro F1 score of 0.90, which is a substantial enhancement from the baseline score of 0.66. The performance of FastGTN was inferior in both the datasets since the target node types were textually attributed and FastGTN though better in identifying patterns from graph topology but was not as good as LLM in making use of the textual component for the prediction task as node classification.

To investigate the contribution of the LLM and the proposed methodology, we conducted an experiment where the LLM was provided with only examples as context. In this scenario, the micro F1 scores were approximately 0.76 for DBLP and 0.77 for Goodreads. These results indicate that the performance gain observed in our proposed methodology is primarily attributed to the incorporation of context from the graph, rather than relying solely on the LLM’s internal memory.

Furthermore, our analysis revealed that the performance improvement when utilizing a heterogeneous context in comparison to that of a homogeneous context can be attributed to the use of verbalized metapaths. These metapaths encode the local graph topology surrounding the target node, enabling the LLM to effectively leverage the contextual information that is highly influential for node classification.

Lastly, we observed an additional performance gain when incorporating examples of labels that exhibit semantic similarity to the target node, compared to using random examples. This finding highlights the importance of leveraging semantic relationships between labels in the node classification task.

### Task 2: Explanation Generation Ability

Based on the results presented in the table, the proposed methodology incorporating metapath as well as semantic examples outperforms the homogeneous context approach in terms of both factual correctness and helpfulness ratings. For the DBLP dataset, the percentage of labels with a factual correctness rating of 3 is slightly

higher in the proposed methodology (97.5%) compared to the homogeneous context approach (95.12%). Furthermore, the proposed methodology shows a substantial improvement in the percentage of labels with a helpfulness rating of 3, achieving 30% compared to just 10% for the homogeneous context. Additionally, the percentage of labels with a helpfulness rating of 2 is lower in the proposed methodology (70%) than in the homogeneous context approach (90%), indicating a higher overall helpfulness.

Similarly, for the Goodreads dataset, the proposed methodology also demonstrates superior performance. The factual correctness rating of 3 is marginally higher at 96.15% compared to 94.23% for the homogeneous context. The helpfulness rating of 3 is significantly better in the proposed methodology (38.46%) versus the homogeneous context (19.23%). The percentage of labels with a helpfulness rating of 2 is reduced in the proposed methodology (57.69%) compared to the homogeneous context approach (76.92%).

These results indicate that integrating metapaths and semantic examples into the context not only enhances the factual correctness of the labels but also significantly improves the perceived helpfulness of the explanations, thereby providing a more effective and user-friendly interpretability framework.

## 8. Conclusion

The proposed approach combines Heterogeneous Graph Neural Networks (HGNNs) with Large Language Models (LLMs) through the utilization of verbalized metapaths. This hybrid methodology demonstrates significant enhancements in prediction accuracy and results in interpretability. The incorporation of verbalized metapaths proves advantageous, providing a nuanced understanding of node interactions that traditional methods often overlook. Furthermore, our approach presents an efficient strategy for mitigating the computational demands associated with fine-tuning large-scale language models, facilitating efficient model deployment. Future research endeavors may focus on extending these techniques to other domains and evaluating their performance in real-world applications where both prediction accuracy and interpretability are crucial for decision-making processes.

## 9. Acknowledgement

Lastly, we would like to express our gratitude to DSO Singapore, for the generous grant and anticipated support that aims to make the vision of this research possible.



## References

- [1] Arkadeep Acharya, Brijraj Singh, and Naoyuki Onoe. 2023. LLM Based Generation of Item-Description for Recommendation System. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1204–1207.
- [2] Chirag Agarwal, Owen Queen, Himabindu Lakkaraju, and Marinka Zitnik. 2023. Evaluating explainability for graph neural networks. *Scientific Data* 10, 1 (2023), 144.
- [3] Federico Baldassarre and Hossein Azizpour. 2019. Explainability techniques for graph convolutional networks. *arXiv preprint arXiv:1905.13686* (2019).
- [4] Nadia Burkart and Marco F. Huber. 2020. A Survey on the Explainability of Supervised Machine Learning. *CoRR* abs/2011.07876 (2020). arXiv:2011.07876 <https://arxiv.org/abs/2011.07876>
- [5] Antonia Creswell, Murray Shanahan, and Irina Higgins. 2022. Selection-inference: Exploiting large language models for interpretable logical reasoning. *arXiv preprint arXiv:2205.09712* (2022).
- [6] Yixiang Fang, Yixing Yang, Wenjie Zhang, Xuemin Lin, and Xin Cao. 2020. Effective and efficient community search over large heterogeneous information networks. *Proceedings of the VLDB Endowment* 13, 6 (2020), 854–867.
- [7] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. 2020. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of The Web Conference 2020*. 2331–2341.
- [8] Thorben Funke, Megha Khosla, and Avishek Anand. 2020. Hard masking for explaining graph neural networks. (2020).
- [9] Huiting Hong, Hantao Guo, Yucheng Lin, Xiaoqing Yang, Zang Li, and Jieping Ye. 2020. An attention-based graph neural network for heterogeneous structural learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 4132–4139.
- [10] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, and Yi Chang. 2022. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering* (2022).
- [11] Rana Hussein, Dingqi Yang, and Philippe Cudré-Mauroux. 2018. Are meta-paths necessary? Revisiting heterogeneous graph embeddings. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 437–446.
- [12] Yanbin Jiang, Huifang Ma, Xiaohui Zhang, Zhixin Li, and Liang Chang. 2024. Incorporating metapath interaction on heterogeneous information network for social recommendation. *Frontiers of Computer Science* 18, 1 (2024), 181302.
- [13] Bowen Jin, Yu Zhang, Qi Zhu, and Jiawei Han. 2023. Heterformer: Transformer-based deep node representation learning on heterogeneous text-rich networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1020–1031.
- [14] Jaykumar Kakkad, Jaspal Jannu, Kartik Sharma, Charu Aggarwal, and Sourav Medya. 2023. A survey on explainability of graph neural networks. *arXiv preprint arXiv:2306.01958* (2023).
- [15] Peibo Li, Yixing Yang, Maurice Pagnucco, and Yang Song. 2022. Explainability in graph neural networks: An experimental survey. *arXiv preprint arXiv:2203.09258* (2022).
- [16] Antonio Longa, Steve Azzolin, Gabriele Santin, Giulia Cencetti, Pietro Liò, Bruno Lepri, and Andrea Passerini. 2022. Explaining the Explainers in Graph Neural Networks: a Comparative Study. *arXiv preprint arXiv:2210.15304* (2022).
- [17] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. 2020. Parameterized explainer for graph neural network. *Advances in neural information processing systems* 33 (2020), 19620–19631.
- [18] Steven Moore<sup>1</sup>, Richard Tong, Anjali Singh, Zitao Liu, Xianguan Hu, Yu Lu, Joleen Liang, Chen Cao, Hassan Khosravi, Paul Denny, et al. 2023. Check for updates Empowering Education with LLMs-The Next-Gen Interface and Content Generation. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky: 24th International Conference, AIED 2023, Tokyo, Japan, July 3–7, 2023, Proceedings*. Springer Nature, 32.
- [19] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.
- [20] Yitong Pang, Lingfei Wu, Qi Shen, Yiming Zhang, Zhihua Wei, Fangli Xu, Ethan Chang, Bo Long, and Jian Pei. 2022. Heterogeneous global graph neural networks for personalized session-based recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 775–783.
- [21] Bhargavi Paranjape, Scott Lundberg, Sameer Singh, Hannaneh Hajishirzi, Luke Zettlemoyer, and Marco Tulio Ribeiro. 2023. ART: Automatic multi-step reasoning and tool-use for large language models. *arXiv preprint arXiv:2303.09014* (2023).
- [22] Phillip E Pope, Soheil Kolouri, Mohammad Rostami, Charles E Martin, and Heiko Hoffmann. 2019. Explainability methods for graph convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10772–10781.
- [23] Thomas Schnake, Oliver Eberle, Jonas Lederer, Shinichi Nakajima, Kristof T Schütt, Klaus-Robert Müller, and Grégoire Montavon. 2021. Higher-order explanations of graph neural networks via relevant walks. *IEEE transactions on pattern analysis and machine intelligence* 44, 11 (2021), 7581–7596.
- [24] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2016. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2016), 17–37.
- [25] Chandan Singh, Jeevana Priya Inala, Michel Galley, Rich Caruana, and Jianfeng Gao. 2024. Rethinking Interpretability in the Era of Large Language Models. *arXiv preprint arXiv:2402.01761* (2024).
- [26] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. 2014. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806* (2014).
- [27] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*. PMLR, 3319–3328.
- [28] Yiming Tan, Dehai Min, Yu Li, Wenbo Li, Nan Hu, Yongrui Chen, and Guilin Qi. 2023. Evaluation of ChatGPT as a question answering system for answering complex questions. *arXiv preprint arXiv:2303.07992* (2023).
- [29] Cunchao Tu, Xiangkai Zeng, Hao Wang, Zhengyan Zhang, Zhiyuan Liu, Maosong Sun, Bo Zhang, and Leyu Lin. 2018. A unified framework for community detection and network representation learning. *IEEE Transactions on Knowledge and Data Engineering* 31, 6 (2018), 1051–1065.
- [30] Minh Vu and My T Thai. 2020. Pgm-explainer: Probabilistic graphical model explanations for graph neural networks. *Advances in neural information processing systems* 33 (2020), 12225–12235.
- [31] Mengting Wan, Rishabh Misra, Ndapa Nakashole, and Julian J. McAuley. 2019. Fine-Grained Spoiler Detection from Large-Scale Review Corpora. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28– August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 2605–2610. <https://doi.org/10.18653/V1/P19-1248>

- [32] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The world wide web conference*. 2022–2032.
- [33] Huance Xu, Chao Huang, Yong Xu, Lianghao Xia, Hao Xing, and Dawei Yin. 2020. Global context enhanced social recommendation with hierarchical graph neural networks. In *2020 IEEE international conference on data mining (ICDM)*. IEEE, 701–710.
- [34] Siyong Xu, Cheng Yang, Chuan Shi, Yuan Fang, Yuxin Guo, Tianchi Yang, Luhao Zhang, and Maodi Hu. 2021. Topic-aware heterogeneous graph neural network for link prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2261–2270.
- [35] Hao Yan, Chaozhuo Li, Ruosong Long, Chao Yan, Jianan Zhao, Wenwen Zhuang, Jun Yin, Peiyan Zhang, Weihao Han, Hao Sun, Weiwei Deng, Qi Zhang, Lichao Sun, Xing Xie, and Senzhang Wang. 2023. A Comprehensive Study on Text-attributed Graphs: Benchmarking and Rethinking. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. <https://openreview.net/forum?id=m2mbfoSuJ1>
- [36] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems* 32 (2019).
- [37] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. 2022. Explainability in graph neural networks: A taxonomic survey. *IEEE transactions on pattern analysis and machine intelligence* 45, 5 (2022), 5782–5799.
- [38] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. 2021. On explainability of graph neural networks via subgraph explorations. In *International conference on machine learning*. PMLR, 12241–12252.
- [39] Seongjun Yun, Minbyul Jeong, Sungdong Yoo, Seunghun Lee, S Yi Sean, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2022. Graph Transformer Networks: Learning meta-path graphs to improve GNNs. *Neural Networks* 153 (2022), 104–119.
- [40] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 793–803.
- [41] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. 2023. Benchmarking large language models for news summarization. *arXiv preprint arXiv:2301.13848* (2023).
- [42] Yue Zhang, David Defazio, and Arti Ramesh. 2021. Relex: A model-agnostic relational model explainer. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 1042–1049.
- [43] Jianan Zhao, Xiao Wang, Chuan Shi, Binbin Hu, Guojie Song, and Yanfang Ye. 2021. Heterogeneous graph structure learning for graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4697–4705.
- [44] Jianming Zheng, Fei Cai, Yanxiang Ling, and Honghui Chen. 2020. Heterogeneous graph neural networks to predict what happen next. In *Proceedings of the 28th International Conference on Computational Linguistics*. 328–338.

# Appendix

## Experimental Details

**Table 4.** LLM inference Hyperparameters

| LLM choice | temperature | top_P | max_tokens_to_sample |
|------------|-------------|-------|----------------------|
| Claude 2.1 | 0           | 1     | 500                  |

**Table 5.** Meta-Path Hyperparameters

|                             |    |
|-----------------------------|----|
| Number of type of metapaths | 10 |
| Total metapath instances    | 40 |

**Table 6.** Fast-GTN training hyperparameters

| Dataset    | Num_layers | Dropout | Num_channels | hidden_dim | lr    |
|------------|------------|---------|--------------|------------|-------|
| heightDBLP | 3          | 0.3     | 4            | 128        | 0.01  |
| Goodreads  | 3          | 0.3     | 8            | 128        | 0.001 |

**Table 7.** Hyperparameters in homogeneous context

| Count of neighbours            | 1-hop | 2-hop | 3-hop |
|--------------------------------|-------|-------|-------|
| Homogeneous context upto 1 hop | 40    | 0     | 0     |
| Homogeneous context upto 2 hop | 20    | 20    | 0     |
| Homogeneous context upto 3 hop | 13    | 13    | 14    |

**Table 8.** Dataset: Label Types and Distribution

| Dataset type | Target node type | Labels  | Label distribution  |
|--------------|------------------|---|---------------------|
| DBLP         | paper            | machine_learning, theoretical_computer_science, computer_networking | Equally distributed |
| Goodsreads   | books            | fiction, non-fiction, romance                                       | Equally distributed |

## Prompt Details

Figure 2. Verbalization of metapath: Prompt template:

### Metapath Summarization Prompt

Human: You are a human whose goal is to generate the meaning of different meta-paths in a heterogeneous graph network. You are provided with all the node types and edge-type and the underlying relation between them. Please find the node types enclosed inside the XML tags <node-type>,</node-type>  
<node-type>node-type</node-type>  
Edge-type and their underlying relation are enclosed between XML tags <edge-relation>,</edge-relation>  
<edge-relation>edge-relation</edge-relation>  
Please find one meta-path summarization example below: meta-path-summarization-example-1  
Can you generate the meaning of the following meta-paths enclosed inside XML tags <meta-path>,</meta-path>  
<meta-path>meta-path</meta-path>  
Write the meaning inside xml tags <meaning> </meaning>. Enclose the meaning of every metapath individually into xml tags <meaning>,</meaning>.

Assistant:

**Figure 3.** Augmented Prompt in the template mentioned in Figure 2.

### Prompt After Augmenting

Human: You are a human whose goal is to generate the meaning of different meta-paths in a heterogeneous graph network. You are provided with all the node types and edge types and the underlying relation between them.

Please find the node types enclosed inside the XML tags <node-type>,</node-type>

<node-type>

P: paper

A: author

C: conference

</node-type>

Edge-type and their underlying relation are enclosed between XML tags <edge-relation>,</edge-relation>

<edge-relation>

PC:=PC (P stands for bcd, S stands for efg) : Paper bcd is published in conference efg

CP:=CP (C stands for efg, P stands for bcd) : Conference efg has paper bcd as one of its submission

PA:=PA (P stands for bcd, A stands for hij) : Paper bcd has been authored by hij

AP:=AP (A stands for hij, P stands for bcd) : Author hij has authored paper bcd

PP:=PP (P stands for bcd, P stands for klm) : Paper bcd has referenced another paper klm

</edge-relation>

Please find one meta-path summarization example below:

BAB (B stands for xyz, A stands for abc, B; stands for def) : Book xyz is related to book def as they are both written by the same author abc. They might be similar semantically or have a similar theme

Can you generate the meaning of the following meta-paths enclosed inside XML tags <meta-path>,</meta-path>

<meta-path>

PAP; (P stands for NbrnT, A stands for P3fAb, P; stands for nFbm0)

PP;A (P stands for NbrnT, P; stands for P3fAb, A stands for nFbm0)

PP;C (P stands for NbrnT, P; stands for P3fAb, C stands for nFbm0)

PA (P stands for NbrnT, A stands for P3fAb)

PC (P stands for NbrnT, C stands for P3fAb)

PP;P;; (P stands for NbrnT, P; stands for P3fAb, P;; stands for nFbm0)

PP; (P stands for NbrnT, P; stands for P3fAb)

PCP; (P stands for NbrnT, C stands for P3fAb, P; stands for nFbm0)

PP;AP;; (P stands for NbrnT, P; stands for P3fAb, A stands for nFbm0, P;; stands for HnKYa)

</meta-path>

Write the meaning inside xml tags <meaning> </meaning>. Enclose the meaning of every metapath individually into XML tags <meaning>,</meaning>.

Assistant:

**Figure 4.** Example generation: Verbalization of metapath

Output Example for Summarizing Meta-path

<meaning>PAP; (P stands for NbrnT, A stands for P3fAb, P; stands for nFbm0): Paper NbrnT is related to paper nFbm0 through author P3fAb. This meta-path suggests that both papers are likely to share thematic or research similarities due to the common authorship.</meaning>

<meaning>PP;A (P stands for NbrnT, P; stands for P3fAb, A stands for nFbm0): Paper NbrnT references paper P3fAb, which is authored by nFbm0. This path could indicate that the research in paper NbrnT is influenced or builds upon the work done by author nFbm0 in paper P3fAb.</meaning>

<meaning>PP;C (P stands for NbrnT, P; stands for P3fAb, C stands for nFbm0): Paper NbrnT references paper P3fAb, which is published in conference nFbm0. This meta-path might suggest a thematic or subject matter connection between the two papers, focused on the themes of the conference.</meaning>

<meaning>PA (P stands for NbrnT, A stands for P3fAb): Paper NbrnT has been authored by P3fAb. This indicates a direct authorial relationship where P3fAb contributes to the content and research findings of the paper NbrnT.</meaning>

<meaning>PC (P stands for NbrnT, C stands for P3fAb): Paper NbrnT is published in conference P3fAb. This meta-path suggests that the paper's subject is relevant to the themes or focus areas of the conference.</meaning>

<meaning>PP;P;; (P stands for NbrnT, P; stands for P3fAb, P;; stands for nFbm0): Paper NbrnT references paper P3fAb, which in turn references another paper nFbm0. This extended chain of references may indicate a progression or evolution of ideas and research across the three papers.</meaning>

<meaning>PP; (P stands for NbrnT, P; stands for P3fAb): Paper NbrnT references paper P3fAb. This relationship implies a direct citation or intellectual linkage, suggesting that NbrnT builds upon or discusses the research presented in P3fAb.</meaning>

<meaning>PCP; (P stands for NbrnT, C stands for P3fAb, P; stands for nFbm0): Paper NbrnT is published in conference P3fAb, and conference P3fAb features another paper nFbm0. This suggests a thematic linkage between the two papers based on their common selection for the same conference.</meaning>

<meaning>PP;AP;; (P stands for NbrnT, P; stands for P3fAb, A stands for nFbm0, P;; stands for HnKYa): Paper NbrnT references paper P3fAb, which is authored by nFbm0 who also authored paper HnKYa. This complex meta-path suggests a deep thematic or research linkage potentially due to the repeated involvement of author nFbm0.</meaning>

Figure 5. Prompt template for getting prediction and explanation

**LLM Only - Get Prediction and Explanation**

Human: {task\_description}  
 -provide the reasoning logic or methodology to get the prediction for the node. Post obtaining the prediction, you have to provide the justification and explanation.

The name of the node for which you have to predict the label and provide an explanation is enclosed in XML tags <node-name>{node\_name}</node-name> and its description is enclosed in <node-description>{node\_description}</node-description>

Please find the examples related to different label types, as enclosed in XML tags <example-context>{examples}</example-context>

Assistant: Thanks for providing the task-description and examples. Do you have any other instructions?

Human:  
 0. Create short notes for the context and the examples in a way that you can relate the description of the node you are required to predict to that of context.  
 1. Discuss the relevance of the given node and its description to the context and examples and write it inside <relevance></relevance> XML tags. This is also a space for you to write down your reasoning and is not meant for the user.  
 2. Provide the methodology to get the prediction for the label of the node. Enclose the methodology in xml tags <methodology></methodology>. This is just the space for your notes and is not meant to be shown to the user. Please make sure this methodology is written in the form of steps.  
 3. Get the prediction of the label of the node, and enclose the prediction in XML tags. <prediction></prediction>  
 4. Get the explanation behind the prediction for the label of the node. Enclose it in XML tags <explanation></explanation>

Assistant:

**Proposed Prompt Template in Proposed Methodology**

Human: {task\_description}  
 -provide the reasoning logic or methodology to get the prediction for the node. Post obtaining the prediction, you have to provide the justification and explanation.

The name of the node for which you have to predict the label and provide an explanation is enclosed in XML tags <node-name>{node\_name}</node-name> and its description is enclosed in <node-description>{node\_description}</node-description>  
 Please find the context for this node enclosed in XML tags <relevant-context>, </relevant-context>:  
 <context>{context}</context>

The above context can consist of various documents having structural as well as textual information related to the connectivity of the node with its neighbors in the graph.

Please find the examples related to different label types, as enclosed in XML tags <example-context>{examples}</example-context>

Assistant: Thanks for providing the task-description, context, and examples. Do you have any other instructions?

Human:  
 0. Create short notes for the context and the examples in a way that you can relate the description of the node you are required to predict to that of context. Please also consider the relationship of {node\_name} to its neighboring nodes while writing your notes.  
 1. Discuss the relevance of the given node and its description to the context and examples and its relationship to other nodes and write it in inside <relevance></relevance> XML tags. This is also a space for you to write down your reasoning and is not meant for the user.  
 2. Provide the methodology to get the prediction for the label of the node. Enclose the methodology in xml tags <methodology></methodology>. This is just the space for your notes and is not meant to be shown to the user. Please make sure this methodology is written in the form of steps.  
 3. Get the prediction of the label of the node, and enclose the prediction in XML tags. <prediction></prediction>  
 4. Get the explanation behind the prediction for the label of the node. The relationship of {node\_name} to its nodes in its neighborhood can help you get a better explanation. Enclose it in XML tags <explanation></explanation>

Assistant:

**Figure 6.** Example Placeholder for Task Description in Proposed Methodology

**Task Description - DBLP Dataset (Field of Study Classification)**

You are provided with an academic citation network and you have to predict the label of the target node and provide an explanation. By target node, we mean to refer to the paper, and by label, we refer to the field of the study of the paper. Please predict the labels from only these categories: <label1>machine\_learning</label1>, <label2>computer\_networking</label2>, <label3>theoretical\_computer\_science</label3>.

**Figure 7.** Description of placeholder for context in proposed methodology

**Meta-path context format**

```

\{metapath instance meaning 1\}\n
info/details related to \{node 1 in metapath instance 1\}: \{key 1\} = \{value 1\}\n
info/details related to \{node 1 in metapath instance 1\}: \{key 2\} = \{value 2\}\n
... (continue for as many key-value pairs as needed for this node)\n
info/details related to \{node 2 in metapath instance 1\}: \{key 1\} = \{value 1\}\n
info/details related to \{node 2 in metapath instance 1\}: \{key 2\} = \{value 2\}\n
... (continue for as many key-value pairs as needed for this node)\n
... (continue for as many nodes as needed in this metapath instance)\n\n
\{metapath instance meaning 2\}\n
info/details related to \{node 1 in metapath instance 2\}: \{key 1\} = \{value 1\}\n
info/details related to \{node 1 in metapath instance 2\}: \{key 2\} = \{value 2\}\n
... (continue for as many key-value pairs as needed for this node)\n
... (continue for as many nodes as needed in this metapath instance)\n\n
... (continue for as many metapath instances as needed)
    
```