Applied Artificial Intelligence Research

INNOVIAIR

Constructing Multimodal Wireless Knowledge Graphs for Large Language Model–Based Network Reasoning

Wenhao Lin[a], Yingjie Song[a], Yinguo Liu[a], Ying Li[a,*]

[a]College of Computer Science and Technology, Qingdao University, Qingdao, Shandong 266071, China

ARTICLEINFO

Keywords:
Knowledge Graph
Retrieval-Augmented Generation
Wireless Networks
Large Language Models
Graph-Based Retrieval

ABSTRACT

The increasing complexity of wireless networks and the proliferation of heterogeneous data sources pose significant challenges to intelligent network management and decision-making. Although Large Language Models (LLMs) have demonstrated strong capabilities in natural language understanding and reasoning, their direct application in wireless network tasks is limited due to the lack of structured domain knowledge and reliable contextual foundations. This paper proposes an intelligent knowledge graph enhanced retrieval-enhanced generation (KG-RAG) framework for multimodal wireless networks. First, through structured entity and relationship modeling, multimodal wireless data including network measurements, configuration parameters and domain knowledge are transformed into a unified knowledge graph. The constructed knowledge graph, as a clear semantic backbone, supports efficient knowledge retrieval and reasoning. By integrating graph-based retrieval results with LLMs, the proposed framework achieves context-aware and interpretable network reasoning. The effectiveness of this method is evaluated through a case study based on lightweight network slicing. The experimental results show that in different LLMs, compared with the traditional text-based RAG method, the KG-enhanced RAG method improves the reasoning accuracy, reasoning consistency and interpretive integrity. These findings indicate that structured knowledge representation plays a crucial role in enhancing the reliability and interpretability of LLM-driven wireless network intelligence.

1.Introduction

Wireless communication systems are developing rapidly towards 5G-Advanced and even more advanced directions, greatly increasing the complexity of network architecture and operation processes[1]. Modern wireless networks feature large-scale deployment, heterogeneous access technologies, dense topological structures and highly dynamic operating environments, similar to urban systems where intelligent prediction tasks[2] have proven critical for data-driven decision-making. Therefore, network management and optimization increasingly rely on the massive heterogeneous information generated by multiple planes of the network, including protocol specifications, configuration files, signaling logs, performance metrics, topology descriptions, etc. The effective organization and utilization of this multimodal wireless network information has become a key challenge for intelligent communication systems[3].

In recent years, artificial intelligence technology has been widely introduced into wireless networks to support tasks such as fault diagnosis, performance optimization and resource management. Data-driven approaches, including deep learning and reinforcement learning, have demonstrated promising results in specific scenarios. However, these methods usually operate like a black box, requiring a large amount of labeled data, which limits their generalization ability and interpretability in complex network environments. Furthermore, most existing methods regard heterogeneous network data as independent inputs and lack a unified semantic representation to capture the intrinsic relationships among network entities, configurations, and operational states.

Recently, large language models (LLMs) have attracted increasing attention in wireless network research, as they exhibit strong capabilities in language understanding, reasoning, and knowledge integration for complex system analysis[4]. To enhance the domain adaptability of LLMs, recent studies have explored integrating external knowledge sources into LLM-based systems, enabling more reliable and

* Corresponding author.
E-mail address: yingli2016@qdu.edu.cn.
https://doi.org/10.65455/1bt00146
Received 1 April 2026; Received in revised form 20 April 2026; Accepted 27 April 2026; Available online 20 May 2026

context-aware reasoning. In particular, retrieval-augmented generation (RAG) has emerged as an effective paradigm to ground LLM outputs on retrieved knowledge, thereby mitigating hallucination and improving factual consistency. However, most existing RAG approaches primarily rely on unstructured text retrieval, which is insufficient for capturing the rich relationships and constraints inherent in wireless network systems[5].

Wireless network knowledge is essentially multimodal and multi-dimensional. The text descriptions in standards and manuals define protocol behavior and constraints, while digital performance metrics reflect the real-time network status. Network logs and alerts record time events, while topological data records the structural dependencies among network components. Text-based retrieval alone often fails to represent such structured dependencies and multimodal characteristics, leading to incomplete or inconsistent reasoning results when LLMs are applied to networking tasks[6]. Therefore, when applied to complex wireless network problems, LLM-based reasoning may be affected by irrelevant context selection, redundant information, and limited interpretability.

Knowledge graphs provide a natural and effective approach for modeling complex systems by representing entities and their relationships in a structured and machine-interpretable form[7]. In a wireless network environment, knowledge graphs can explicitly encode network components, protocol elements, performance metrics, and operational events, as well as their semantic and functional dependencies. By integrating heterogeneous data sources into a unified knowledge representation, knowledge graphs can achieve structured queries, relationship-aware reasoning, and incremental knowledge evolution. These features make knowledge graphs a promising foundation for enhancing LLM-based reasoning in intelligent communication systems.

Based on these observations, this paper proposes a unified framework as shown in Figure 1 to construct a wireless network knowledge graph from multimodal network information and utilize it to achieve knowledge-enhanced retrieval - enhanced generation for intelligent network reasoning. This method does not directly retrieve unstructured text, but retrieves subgraphs related to semantics and structure that reflect the underlying network logic. Then, these subgraphs are transformed into structured context representations to guide the LLMs in performing netent-aware reasoning tasks, such as fault analysis, performance diagnosis, and configuration decision support.

The key idea of this work is to position the knowledge graph as the intermediate semantic layer between multimodal wireless network data and LLMs. This design Bridges the gap between data-driven intelligence and knowledge-driven reasoning, allowing LLMs to operate in web-related contexts rather than isolated text fragments. By integrating topological information, protocol dependencies and operational relationships into the retrieval process, the proposed framework enhances the relevance, consistency and interpretability of LLM-generated responses in wireless network scenarios.

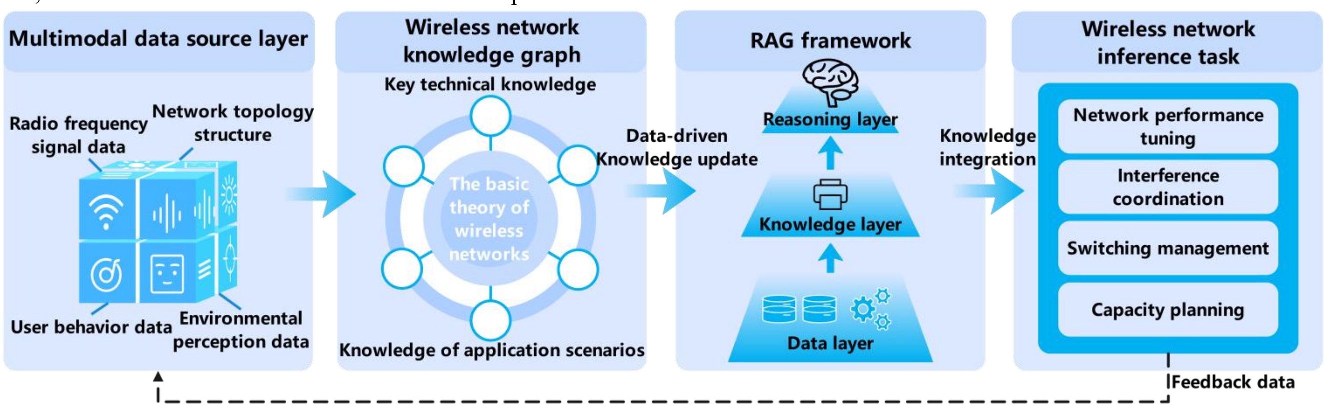

Fig 1. Knowledge-enhanced reasoning framework for intelligent wireless networks

The main contributions of this article can be summarized as follows:

(1)We propose a multimodal semantic alignment framework for wireless networks, which unifies heterogeneous data sources—including protocol descriptions, topology structures, operation logs, and performance metrics—into a consistent knowledge representation. Unlike traditional wireless knowledge graphs that rely on single-modality or static data, this approach explicitly captures cross-modal semantic dependencies and enables coherent integration of heterogeneous network information.

(2)We design a dynamic knowledge graph construction mechanism that incorporates temporal evolution and event-driven updates. By modeling network events, temporal dependencies, and performance variations as first-class knowledge components, the proposed method supports continuous knowledge fusion and reflects the dynamic nature of wireless networks, which is not addressed in conventional static knowledge graph approaches.

(3)We develop a wireless network–oriented graph-based retrieval-augmented generation framework, where graph-aware retrieval preserves topological dependencies, protocol constraints, and causal relationships. Compared with traditional text-based RAG and generic KG-RAG methods, the proposed framework enables structure-aware and network-consistent reasoning, significantly improving reasoning accuracy, consistency, and interpretability in wireless network scenarios.

The rest of this article is organized as follows. The second section reviews the relevant work in wireless network knowledge modeling, intelligent network management, and LLM-based reasoning. The third part analyzes the characteristics and challenges faced by multimodal wireless network information. The fourth part describes the construction of the knowledge graph for wireless networks. Section Five introduces the knowledge-enhancing RAG framework for intelligent network reasoning. Section 6 provides case studies and evaluation results. Finally, the seventh part summarizes the full text and looks forward to the future research directions.

2.Related work

This section reviews existing studies related to our work from three perspectives: knowledge modeling in wireless networks, intelligent network management and optimization, and the application of LLMs with retrieval-augmented generation in communication systems.

2.1.Knowledge modeling in wireless networks

Knowledge modeling has always been regarded as an effective method for representing complex communication systems and supporting intelligent network management. Early research mainly focused on ontology-based modeling of network components, protocol layers and service behaviors, aiming to provide a unified semantic representation for network configuration and management[8]. These methods have been applied in fields such as policy-based network control, fault management and service orchestration.

With the increasing complexity of wireless networks, knowledge graph technology has been introduced to capture the richer relationships among network entities. To systematically organize such complex and interrelated information, knowledge graphs have been explored as a structured representation for wireless network systems. Recent work has explored the construction of wireless network knowledge graphs from heterogeneous sources such as standard documents, configuration files, and operation logs. These knowledge graphs typically encode entities such as base stations, users, links, and protocols, as well as relationships that describe topology, dependencies, and functional interactions. Multimodal wireless knowledge graphs have been shown to enhance the expressiveness of network knowledge modeling and provide a foundation for advanced learning and reasoning tasks[9]. Knowledge graphs show potential in supporting network diagnosis and optimization tasks by supporting structured queries and reasoning.

However, most of the existing knowledge graphs of wireless networks are manually constructed or rely heavily on rule-based extraction methods, which limits their scalability and adaptability to dynamic network environments. Furthermore, many studies focus on static knowledge representation and lack effective mechanisms to combine real-time operational data with temporal evolution. More importantly, existing knowledge modeling efforts are typically designed for traditional queries or rule-based reasoning rather than for integration with data-driven or LLM-based reasoning frameworks.

2.2.Intelligent network management and optimization

Artificial intelligence has been widely applied in wireless networks to address challenges in network management and optimization. Existing studies have demonstrated that machine learning techniques, including supervised learning, unsupervised learning, and reinforcement learning, can effectively support traffic prediction, resource allocation, switching optimization, and fault detection in wireless systems[10]. Similar advances have been achieved in cross-domain intelligent prediction tasks: travel demand prediction integrating natural environmental and socioeconomic factors[11], deep learning-based urban traffic revitalization index modeling during the COVID-19 epidemic[12], and deep spatial-temporal travel time prediction leveraging trajectory features[13], and advanced shared mobility demand forecasting models that integrate tensor decomposition and spatio-temporal recurrent networks to handle data sparsity and complex dependencies[14], road segmentation for remote sensing images based on dedicated deep learning networks[15]. These methods utilize a large amount of network data to learn implicit patterns and decision-making strategies. Similar approaches have been widely explored in spatial-temporal prediction tasks, such as taxi demand forecasting based on spatial-temporal graph optimization models, which effectively capture dynamic dependencies in complex systems[16].

In recent years, there has been an increasing focus on combining domain knowledge with data-driven learning to enhance the robustness and interpretability of models. Hybrid approaches that integrate expert knowledge, constraints, or network models into learning frameworks have been proposed to achieve more informed and explainable decision-making[17]. For instance, knowledge-assisted reinforcement learning and model-based learning methods have been proposed to leverage network structures and protocol logic.

Despite these advancements, intelligent network management methods still face some limitations. Many methods require a large amount of training data and meticulous parameter tuning, which makes them difficult to be generalized across different network deployments. Furthermore, most learning-based solutions focus on specific tasks and lack a unified knowledge representation that can be reused across multiple network functions. As a result, network intelligence remains fragmented, and cross-domain reasoning involving topological, configuration, and performance information remains challenging[18].

2.3.Large language models and RAG in communication systems

In recent years, LLMs have attracted much attention due to their powerful capabilities in natural language understanding and reasoning. In the field of networking, preliminary research has explored the application of LLMs in tasks such as network troubleshooting, configuration suggestions, and technical document question answering. Recent studies have reported promising results when applying LLMs to assist network operation and management tasks. To alleviate the limitations of LLMs in terms of knowledge and factual accuracy in specific fields, retrieval enhanced generation was introduced, incorporating external information sources into the reasoning process.

The existing RAG-based methods in communication systems usually retrieve relevant documents, logs or configuration files based on semantic similarity and provide them as context input to LLMs. Compared with independent LLMs, these methods have improved performance in answering network-related queries. However, most current implementations rely on unstructured text retrieval and simple block-based indexing, which has been shown to exhibit limitations in preserving relationships and structural dependencies in complex domains[19].

Wireless networks involve complex relationships among network elements, protocols and performance metrics, which cannot be fully captured by planar text representation. Recent research has shown that graph-based RAG methods can better preserve structural dependencies and enhance retrieval effectiveness for structured reasoning tasks[20]. Furthermore, the lack of a clear structure limits the interpretability of the responses generated by LLMs, which is a key requirement for network management and decision support.

In conclusion, the existing research on wireless network knowledge modeling provides valuable insights into the structured representation of network information, but to a large extent, it is still isolated from modern LLM-based reasoning frameworks. Intelligent network management methods have achieved remarkable success by leveraging data-driven approaches, but they often lack unified knowledge representation and cross-domain reasoning capabilities. Recent attempts to apply LLMs and RAG in communication systems demonstrate strong potential, but their reliance on unstructured text retrieval constrains reasoning reliability and interpretability[21].

These limitations highlight the need for a unified framework that integrates multimodal wireless network information into structured knowledge representations and utilizes it to enhance LLM-based reasoning. By combining knowledge graph modeling with retrieval-augmented generation, recent studies suggest that the gap between heterogeneous network data and intelligent reasoning can be effectively bridged[22], thereby providing more accurate, consistent and interpretable decision support for wireless networks.

3.Multimodal wireless network information modeling

The operation and management of modern wireless networks rely on diverse information sets generated across multiple network planes and functional modules. In traditional communication systems, network status can be characterized by a limited number of parameters, while modern wireless networks generate heterogeneous information with different representations, time granularities and semantic meanings. Recent studies on beyond-5G and 6G networks have emphasized that such heterogeneity fundamentally changes how network intelligence should be designed and supported[23]. Effective modeling of this multimodal wireless network information is a prerequisite for achieving intelligent network reasoning and decision support.

3.1.Multimodal information sources in wireless networks

Wireless network information comes from multiple sources, and each source reflects different aspects of network behavior and operation.

Protocol and specification information: Standard documents, protocol descriptions, and technical manuals define the functional logic, signaling processes, and configuration constraints of wireless systems. This type of information is typically represented in natural language and formal definitions, providing advanced semantic knowledge about network behavior, protocol interaction, and design principles. Protocol knowledge has been widely recognized as an essential source of semantic context for intelligent network management and automation. Although protocol knowledge is relatively static, it plays a crucial role in interpreting network events and guiding decision-making.

Configuration and topology information: Configuration files and topology descriptions specify the interconnection and parameterization methods of network components such as base stations, core network elements, and links. This information explicitly captures the structural features of the network, including hierarchical relationships, dependency constraints, and connection patterns. Prior work has shown that topological and configuration information is critical for root-cause analysis and dependency-aware network reasoning[24]. Topological and configuration data form the backbone of network-level inference, as many operational issues are closely related to structural dependencies.

Operation logs and signaling data: Network logs, alerts, and signaling tracking record detailed operation events during network operation. These data sources typically come with timestamps and event-driven features, reflecting dynamic network behaviors such as handovers, failures, and protocol state transitions. Event logs and signaling traces have been extensively used to characterize network dynamics and diagnose transient failures[25]. Operation logs provide valuable insights into the temporal evolution of network status, but are essentially unstructured or semi-structured.

Performance indicators and monitoring data: Key performance indicators (KPIs) quantitatively describe the performance status of the network, such as throughput, latency, packet loss, signal quality, etc. These indicators are usually collected in the form of time series, reflecting real-time or near real-time network conditions. Time-series performance data has become a core input for data-driven network optimization and anomaly detection methods[26]. Performance data is crucial for detecting anomalies, diagnosing performance degradation, and evaluating optimization strategies.

These information sources together form a highly heterogeneous and multimodal information space, posing significant challenges to unified modeling and intelligent utilization.

3.2.Characteristics and challenges of multimodal wireless information

Multimodal wireless network information exhibits several inherent characteristics, complicates its direct application in intelligent reasoning systems.

The heterogeneity of representation is a fundamental challenge. Protocol knowledge is mainly textual and semantic, topological information is graphically structured, logs are event-based, and performance metrics are numerical and temporal. Several recent surveys on network intelligence have highlighted representation heterogeneity as a major obstacle to unified modeling and reasoning[27]. These heterogeneous representations make it very difficult to apply a single modeling or retrieval strategy without losing important information.

Cross-modal semantic dependencies are very common in wireless networks. For instance, the performance degradation observed in KPI data might be caused by a series of signaling events, which are controlled by protocol logic and constrained by the network topology. Such cross-modal dependencies have been identified as a key challenge for isolated data processing and task-specific learning approaches[28]. This cross-modal dependency cannot be captured through isolated data processing or simple feature joining.

Time dynamics and evolution play a crucial role in the operation of networks. Capturing such temporal patterns has been extensively studied in time-series forecasting tasks, where hybrid approaches combining signal decomposition and deep learning are employed to model complex and non-stationary dynamics[29]. Due to changes in traffic patterns, mobility and environmental conditions, wireless networks are constantly evolving. Logs and performance metrics reflect temporary states, while topologies and configurations may also change over time. Modeling temporal evolution has been shown to be essential for accurate fault diagnosis and proactive network optimization[30].

The scale and complexity further intensify the challenges of modeling. Large-scale wireless networks may consist of thousands of network elements and generate vast amounts of data. Under such scale constraints, it is very important to effectively organize and access relevant information, especially when real-time or near real-time reasoning is required.

3.3.Limitations of direct data-driven and text-based approaches

Existing intelligent network solutions often rely on direct data-driven learning or text-based information retrieval. Although these methods have achieved success in specific tasks, they exhibit inherent limitations when applied to multimodal wireless network information.

Pure data-driven learning methods usually require a large amount of training data and are highly task-specific, and their effectiveness often depends on the integration of multiple influencing factors and complex interaction mechanisms, as observed in empirical studies involving multi-factor modeling and moderating effects[31]. They usually lack explicit representations of network structure and protocol logic, thus making it difficult to generalize across scenarios or provide interpretable reasoning results. Recent studies have pointed out that the absence of structural knowledge limits the robustness and explainability of purely data-driven network intelligence solutions.

Text-based retrieval and RAG methods treat network information as unstructured text and retrieve relevant content based on semantic similarity. However, this method cannot maintain the structural and relational nature of wireless network information, a limitation also observed in multi-source heterogeneous data fusion and time-series prediction tasks such as traffic forecasting using advanced neural architectures[32,33], multi-feature attention graph convolutional network for traffic prediction[34], and fast autoregressive tensor decomposition for real-time traffic flow prediction[35], where unstructured integration fails to capture critical cross-source dependencies. During the process of text blocking, topological constraints, causal relationships, and temporal dependencies are often segmented or lost, which has been recognized as a fundamental limitation of text-centric RAG in structured domains[36]. This leads to incomplete or misleading context inputs for LLMs.

These limitations indicate that neither traditional data-driven learning nor naive text retrieval is sufficient to fully utilize the richness of information in multimodal wireless networks.

To address the above challenges, it is necessary to adopt a modeling paradigm that can unify heterogeneous information, maintain semantic relationships and support structured reasoning. Knowledge-oriented modeling provides a promising solution by abstracting multimodal data into entities, relationships, and attributes that reflect the underlying logic of wireless networks.

By representing network components, protocols, events and performance metrics as interconnected knowledge elements, cross-schema dependencies and temporal evolution can be explicitly captured. This structured representation supports relationship-aware retrieval and reasoning, which is crucial for intelligent network management tasks.

Furthermore, knowledge-oriented models serve as an effective intermediate layer between raw network data and intelligent reasoning engines. LLMs can be guided by structured and semantically coherent contexts derived from web knowledge representations, rather than operating on fragmented or unstructured inputs. This design enhances the relevance and interpretability of the reasoning results.

Based on this motivation, the next section will construct a knowledge graph for wireless networks, integrating multimodal network information into a unified and evolvable knowledge representation, laying the foundation for knowledge-enhanced retrieval - enhanced generation in intelligent wireless networks.

4.Knowledge graph construction for wireless networks

Unlike conventional wireless knowledge graphs, this work emphasizes multimodal semantic alignment and temporal knowledge integration to support dynamic and cross-modal reasoning. In order to achieve structured reasoning on multimodal wireless network information, heterogeneous data sources must be transformed into a unified machine-interpretable representation. In this work, we constructed a knowledge graph for wireless networks. The construction process is shown in Figure 2. This knowledge graph explicitly models network entities, the relationships between entities, and dynamic attributes, serving as the semantic basis for intelligent network reasoning.

4.1.Knowledge schema and entity modeling

The construction of the wireless network knowledge graph is guided by several key design principles:

(1) The knowledge graph should reflect the intrinsic structure of wireless networks. Network elements such as base stations, users, links, and protocols are not independent entities but are connected through well-defined topological, functional, and logical relationships. These relationships must be explicitly represented to support topology-aware and dependency-aware reasoning.

(2) The knowledge graph should integrate multimodal information in a unified manner. Textual protocol knowledge, configuration data, operational events, and performance metrics should be mapped into a common semantic space, enabling cross-modal association and reasoning.

(3) The knowledge graph should support temporal evolution and incremental updates. Wireless networks are highly dynamic, and the knowledge representation must be able to incorporate new events, measurements, and configuration changes without requiring complete reconstruction.

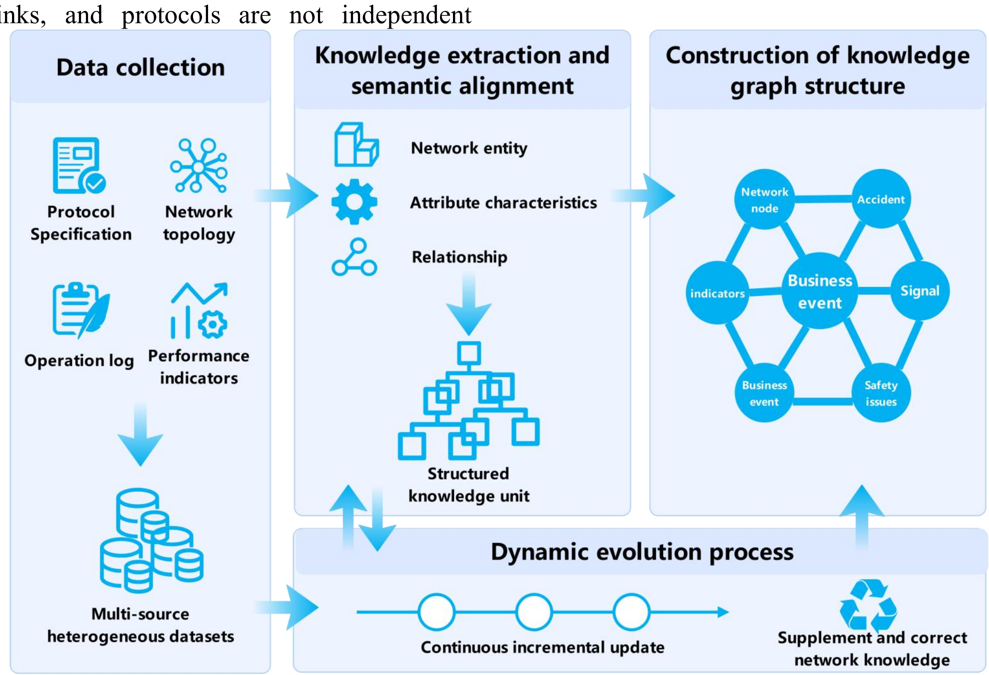

Fig 2. Construction of a multimodal knowledge graph for wireless networks

The constructed knowledge graph should be compatible with retrieval-augmented generation, allowing efficient extraction of semantically and structurally relevant subgraphs as contextual inputs for LLMs.

Based on the characteristics of wireless networks, we define a hierarchical knowledge schema that captures both structural and operational aspects of the network:

(1) Network entities: Core network elements, including base stations, access points, users, and core network functions, are modeled as entities. These entities form the fundamental building blocks of the knowledge graph and represent the physical and logical components of the wireless system.

(2) Protocol and functional entities: Protocol modules, signaling procedures, and functional blocks defined in wireless standards are modeled as abstract entities. These entities encode protocol-level knowledge and define the operational logic governing network behavior.

(3) Event entities: Operational logs, alarms, and signaling traces are mapped to event entities with associated timestamps and contextual attributes. Event entities enable temporal reasoning and capture the dynamic evolution of network states.

(4) Performance and metric entities: Key performance indicators and monitoring metrics are modeled as attribute-rich entities or entity attributes, linking quantitative measurements to corresponding network components and events.

This entity-centric modeling approach allows heterogeneous information to be represented in a consistent and extensible manner.

4.2.Relationship modeling and semantic associations

Relationships play a critical role in expressing the semantics of wireless networks. In the knowledge graph, multiple types of relationships are defined to capture different aspects of network behavior:

(1) Topological relationships, such as connectivity and adjacency, describe how network elements are physically or logically connected.

(2) Functional relationships represent protocol dependencies and operational logic, such as control-flow or signaling interactions.

(3) Causal and temporal relationships link events to their triggering conditions and resulting impacts, enabling cause-effect analysis.

(4) Measurement relationships associate performance metrics with corresponding network components and events.

By explicitly modeling these relationships, the knowledge graph preserves essential network semantics that are often lost in unstructured data representations.

4.3.Multimodal knowledge extraction and integration

Multimodal wireless network information is transformed into a knowledge graph through a unified extraction and integration process.

Text protocol descriptions and documents are processed to extract protocol entities and functional relationships. By parsing configuration files and topology data, network entities and structural connections are identified. Operation logs and signaling data are mapped to event entities with temporal attributes, while performance metrics are associated with relevant network components through metric relationships.

To address the inconsistency and redundancy among different data sources, semantic alignment and entity linking mechanisms were applied. This integration process ensures that the knowledge from multiple modalities contributes consistently to the entire graph structure.

The constructed wireless network knowledge graph provides a structured and semantically rich representation of multimodal network information. By unifying topology, protocol logic, operation events and performance metrics, knowledge graphs support relations-aware retrieval and reasoning.

Importantly, this knowledge representation is designed as an intermediate semantic layer between the raw network data and LLMs. Rather than retrieving fragmented text information, the reasoning engine can operate on graph-based substructures to maintain network semantics and dependencies.

In the next section, we will introduce a knowledge-enhanced retrieval - enhanced generation framework, which utilizes wireless network knowledge graphs to guide LLMs in performing intelligent network inference tasks.

5.Knowledge-enhanced retrieval-augmented generation for intelligent network reasoning

Different from generic KG-RAG frameworks, the proposed approach explicitly incorporates wireless network semantics such as topology, protocol logic, and temporal dependencies into the retrieval process. Although wireless network knowledge graphs provide a structured and unified representation of multimodal network information, an effective mechanism is still needed to utilize this knowledge for intelligent reasoning and decision support. In this section, we will introduce a knowledge-enhanced RAG framework tailored for wireless networks, as shown in Figure 3. This framework integrates graph-aware retrieval and structured context construction to guide LLMs in performing network-centric inference tasks.

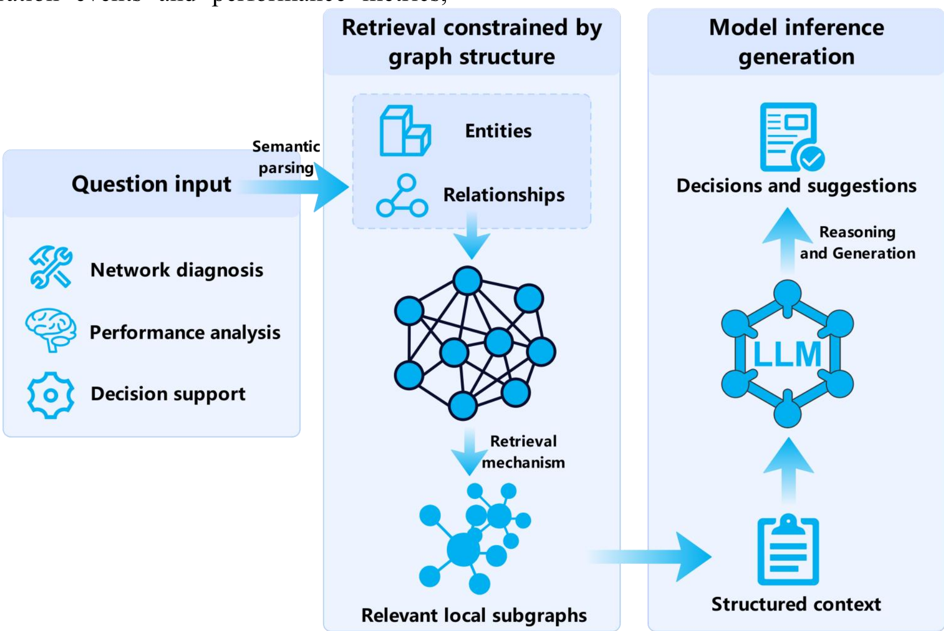

Fig 3. Knowledge graph enhanced RAG framework for intelligent wireless network reasoning

5.1.Overview of the knowledge-enhanced RAG framework

This framework positions the wireless network knowledge graph as the intermediate semantic layer between the raw network data and LLMs. This framework does not directly retrieve unstructured text but retrieves subgraphs related to semantics and structure that reflect the underlying network logic.

As shown in the framework design, network queries or inference requests are first mapped to graph-aware retrieval operations. Then, relevant subgraphs are extracted based on entity correlation, relationship constraints, and temporal context. These subgraphs are then converted into structured text representations, serving as the context input for the LLMs to perform intelligent network inference.

This design enables LLMs to manipulate knowledge related to the network rather than isolated text fragments, thereby enhancing the accuracy, consistency and interpretability of reasoning.

The traditional RAG system mainly relies on the semantic similarity between text embeddings to retrieve relevant documents or text blocks. However, these methods ignore the structural and relational characteristics of wireless networks. To address this limitation, we propose a knowledge-guided retrieval mechanism that explicitly utilizes wireless network knowledge graphs.

Given a network query, relevant entities are first identified through semantic matching or rule-based mapping. Then, the retrieval process unfolds from these entities by traversing the knowledge graph along predefined relationship types such as topological dependencies, functional interactions, and causal associations. Time constraints can also be applied to limit the retrieval within a specific time window.

Through graph traversal and filtering, the retrieval process generates a compact subgraph that captures the most relevant network components, events, and performance metrics associated with the query. Compared with text-based retrieval, this method ensures that the retrieval context retains the network structure and operational semantics.

The retrieved subgraphs provide a structured representation of network knowledge, but LLMs require text input for reasoning. Therefore, an important step of this framework is to convert graph-based knowledge into structured text context.

This method does not linearize the subgraph into plain text, but organizes the extracted knowledge based on network semantics. Entity, relationship and temporal information are presented in a structured and hierarchical manner, retaining topology, causality and protocol logic. For example, describe the network components and their connectivity, followed by related events and related performance indicators.

This structured context construction reduces redundancy and ambiguity, while providing clear network semantics for LLMs. Therefore, the reasoning process becomes more concentrated and consistent with the network behavior in the real world.

5.2.LLM-based network reasoning and decision support

Taking structured context as input, LLMs are adopted to perform network-centric reasoning tasks. Unlike general question responses, network reasoning requires understanding the complex dependencies among topology, protocol behavior, events, and performance states.

Guide the LLMs to analyze the provided network context and generate inference results, such as the cause of the failure, performance bottlenecks, or configuration suggestions. Because the input context explicitly encodes the network relationship and temporal evolution, the generated response is more in line with network logic and easier to interpret.

Importantly, the reasoning results can be traced back to the subgraph of the bottom layer, thereby achieving interpretability and facilitating verification by network operators. This attribute is particularly valuable in actual network management scenarios because transparent decision-making is indispensable.

Compared with the traditional text-based RAG methods, the knowledge-enhanced RAG framework offers several advantages:

(1) It improves retrieval relevance by incorporating structural and relational constraints, reducing the inclusion of irrelevant or redundant context.

(2) It enhances reasoning consistency by preserving network semantics that are often lost in text chunking.

(3) It provides better interpretability through explicit mapping between reasoning outcomes and underlying network knowledge.

These advantages make this framework highly suitable for intelligent wireless network management, as complex dependencies and dynamic behaviors must be considered as a whole.

The knowledge-enhanced RAG framework is designed to be flexible and scalable. Although the current implementation focuses on wireless network scenarios, its basic principles can be applied to other communication systems with complex structures and heterogeneous information sources.

By combining structured knowledge representation with reasoning based on LLMs, this framework takes a step forward towards knowledge-driven intelligent communication systems. It supplements the existing data-driven methods and provides scalable solutions for network inference and decision support in future wireless networks.

6.Case study and evaluation

This section presents a simulation-based case study to quantitatively evaluate the effectiveness of the proposed Knowledge Graph Enhanced Retrieval - Enhanced Generation (KG-RAG) framework in the network slicing scenario. The evaluation is conducted under the heterogeneous service requirements corresponding to enhanced mobile Broadband (eMBB) and ultra-reliable Low Latency Communication (URLLC) films. Compare text-based RAG with graph-based RAG.

6.1.Experimental environment

To verify the feasibility and effectiveness of the wireless network knowledge graph, this experiment constructs a representative wireless communication environment, which covers multiple dimensions such as user service requests, channel state parameters, and spatial location information. These data collectively constitute the original information source of the knowledge graph, providing a foundation for subsequent semantic extraction, entity relationship construction, and reasoning verification.

In modern wireless communication systems, the concept of network slicing has been introduced to meet the performance requirements of different types of services. Network slicing refers to the division of multiple logically independent and functionally customized network "slices" on the same physical network infrastructure through virtualization technology. Each component can provide different network functions for different business scenarios. In short, it divides a network into multiple dedicated networks to achieve the goal of "on-demand service". This mechanism is one of the important features of 5G and future 6G systems, and it is also the core means to achieve multi-service collaboration and resource optimization.

In this study, considering that wireless network knowledge graphs are mainly used to express the correlation between network states and user requirements in highly dynamic and highly heterogeneous scenarios, only two of the most representative network slices were selected: Enhanced Mobile Broadband and Ultra-reliable Low Latency Communication. These two types of slices respectively represent the two ends of the high throughput demand and the low latency demand, and have significant verification value for the expression ability and semantic recognition ability of knowledge graphs.

A heterogeneous wireless network slicing scenario was constructed, which consists of 30 user requests and is distributed in a two-dimensional service area with a fixed terminal height. This setting achieves controllable and realistic evaluation of slice-aware reasoning by jointly considering semantic service description and quantitative network measurement.

Convert multimodal data into structured knowledge graphs. This diagram consists of five types of entities: users, service requests, slice types, network metrics, and locations. Establish relationships to represent request associations, slicing requirements, metric observations, and spatial contexts. By explicitly modeling these relationships, the knowledge graph retains the cross-layer dependencies between service semantics and network conditions. This structured representation enables graph-based retrieval to identify context-related subgraphs and then convert them into structured text input for inference by LLMs.

6.2.Evaluation results

Table 1 presents a quantitative comparison between the text-based RAG method and the proposed graph-based RAG approach. To ensure the reliability of the results, each experiment was repeated multiple times under identical settings. The reported values represent the mean performance along with the corresponding standard deviations.

Table 1. Quantitative comparison between text-RAG and graph-RAG.

Method	Slice Accuracy(%)	Reasoning Consistency	Explanation Completeness(%)
Text-RAG	78.3±2.1	0.62±0.03	64.5±2.8
Graph-RAG	91.7±1.4	0.84±0.02	88.2±1.9

As shown in Table 1, the proposed Graph-RAG method consistently outperforms the text-based RAG across all evaluation metrics. Specifically, the slice intent recognition accuracy improves from 78.3% to 91.7%, achieving an absolute gain of 13.4 percentage points. In addition, the reasoning consistency score increases from 0.62 to 0.84, indicating that the generated results are more aligned with the underlying network conditions and logical dependencies.

Furthermore, the explanation completeness improves significantly, demonstrating that the graph-based retrieval mechanism is more effective in preserving structural relationships and contextual dependencies during the reasoning process. Compared with text-based retrieval, the proposed method provides more coherent and semantically structured context for large language models.

Notably, the relatively small standard deviations across repeated experiments indicate that the proposed framework achieves stable and robust performance under different runs. This further validates the reliability of the knowledge-enhanced retrieval mechanism.

Overall, the results in Table 1 demonstrate that integrating structured knowledge graphs into the RAG framework significantly enhances slice-aware reasoning in heterogeneous wireless network scenarios. By combining semantic service descriptions with quantitative network measurements, the proposed method enables more accurate, consistent, and interpretable network reasoning.

7.Conclusion and future work

This paper studies the effective organization and utilization of heterogeneous multimodal data in wireless networks to support intelligent analysis and decision-making. To address the limitations of traditional text-centered or unstructured retrieval methods, we propose a unified framework to construct multimodal knowledge graphs from different wireless network data sources and integrate them with the Knowledge Enhanced RAG paradigm.

Specifically, a structured knowledge graph was constructed to represent key network entities, events, performance metrics and their relationships, thereby achieving a consistent semantic organization of protocol specifications, topological information, operation logs and monitoring metrics. By leveraging the inherent structural and relational attributes of knowledge graphs, the proposed RAG framework performs graph-guided retrieval to obtain context-relevant subgraphs, which are then transformed into structured inputs for LLMs. This design allows the reasoning process to be constrained by domain knowledge, thereby enhancing the interpretability and relevance of the generated results in wireless network scenarios.

This method highlights the potential of combining knowledge graphs with LLMs for intelligent wireless network management, diagnosis, and optimization. This framework does not view multimodal data as isolated or flat text information, but emphasizes structured knowledge representation and reasoning, which are crucial for complex, dynamic and large-scale communication systems.

Future work will focus on expanding the proposed framework in several directions. More fine-grained temporal and causal models will be incorporated into the knowledge graph to better capture the dynamic behavior of wireless networks. Explore adaptive graph retrieval strategies to further enhance scalability and efficiency in large-scale deployments. The integration of large-scale models and online learning mechanisms in specific fields will be studied to enhance the framework's real-time decision support and autonomous network optimization capabilities.

References

[1] ODIDA M O. The evolution of mobile communication: A comprehensive survey on 5G technology. Journal of Sensor Networks and Data Communications, 2024, 4(1): 1-11.

[2] XU Z, LV Z, CHU B, et al. Progress and prospects of future urban health status prediction. Engineering Applications of Artificial Intelligence, 2024, 129: 107573.

[3] MAO S, LIU L, HOU X, et al. Multi-domain resource management for space–air–ground integrated sensing, communication, and computation networks. IEEE Journal on Selected Areas in Communications, 2024, 42(12): 3380-3394.

[4] NAZAR A M, SELIM M Y, QIAO D, et al. NextG-GPT: Leveraging GenAI for Advancing Wireless Networks and Communication Research//2025 34th International Conference on Computer Communications and Networks (ICCCN). New York: IEEE, 2025: 1-9.

[5] ZHU X, XIE Y, LIU Y, et al. Knowledge graph-guided retrieval augmented generation//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2025: 8912-8924.

[6] MAVROMATIS C, KARYPIS G. Gnn-rag: Graph neural retrieval for large language model reasoning. arXiv preprint, 2024.

[7] YUAN D, ZHOU H, WU D, et al. Enhancing large language models (llms) for telecommunications using knowledge graphs and retrieval-augmented generation//2025 IEEE International Conference on Communications Workshops (ICC Workshops). New York: IEEE, 2025: 486-491.

[8] HUANG Y, YOU X, ZHAN H, et al. Learning wireless data knowledge graph for green intelligent communications: Methodology and experiments. IEEE Transactions on Mobile Computing, 2024, 23(12): 12298-12312.

[9] LIU S, ZHANG T, FU N, et al. Fine-grained graph representation learning for heterogeneous mobile networks with attentive fusion and contrastive learning//Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 39, No. 18. Washington, DC: AAAI Press, 2025: 18933-18942.

[10] DAI Y, LYU L, CHENG N, et al. A survey of graph-based resource management in wireless networks—Part II: Learning approaches. IEEE Transactions on Cognitive Communications and Networking, 2024, 11(4): 2101-2122.

[11] XU Z, LV Z, LI J, et al. A novel perspective on travel demand prediction considering natural environmental and socioeconomic factors. IEEE Intelligent Transportation Systems Magazine, 2022, 15(1): 136-159.

[12] LV Z, LI J, DONG C, et al. Deep learning in the COVID-19 epidemic: A deep model for urban traffic revitalization index. Data & Knowledge Engineering, 2021, 135: 101912.

[13] SHENG Z, LV Z, LI J, et al. Deep spatial-temporal travel time prediction model based on trajectory feature. Computers and Electrical Engineering, 2023, 110: 108868.

[14] YAN H, LI J, CHU B, et al. HT-STNet: a hierarchical Tucker decomposition and spatio-temporal LSTM network for accurate and efficient shared mobility demand forecasting on sparse data. Applied Intelligence, 2025, 55(7): 631.

[15] YUAN G, LI J, LV Z, et al. DDCAttNet: road segmentation network for remote sensing images//International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer International Publishing, 2021: 457-468.

[16] LI J, LV Z, MA Z, et al. Optimization of spatial-temporal graph: A taxi demand forecasting model based on spatial-temporal tree. Information Fusion, 2024, 104: 102178.

[17] ZHAO M, ZHANG Y, LIU Q, et al. AdaSlicing: Adaptive online network slicing under continual network dynamics in open radio access networks//IEEE INFOCOM 2025-IEEE Conference on Computer Communications. New York: IEEE, 2025: 1-10.

[18] LIU Z. Explainable machine learning for telecom customer churn prediction and actionable retention strategies. Applied Artificial Intelligence Research, 2026, 2(1).

[19] ZHOU J. Personalized recommendation algorithms in digital media: A review of engagement effects, information cocoons, and marketing implications. Applied Artificial Intelligence Research, 2026, 2(1).

[20] LI M, MIAO S, LI P. Simple is effective: The roles of graphs and large language models in knowledge-graph-based retrieval-augmented generation. arXiv preprint, 2024.

[21] LIU Z. Explainable machine learning for telecom customer churn prediction and actionable retention strategies. Applied Artificial Intelligence Research, 2026, 2(1).

[22] LI L. Map of geographical science based on artificial intelligence. Applied Artificial Intelligence Research, 2025, 1(3).

[23] NIE H. Ethics-first lightweight AI: Differentiated English teaching and cultural competence in junior middle schools. Applied Artificial Intelligence Research, 2025, 1(3).

[24] LUO H, YAN Y, BIAN Y, et al. Ai reasoning for wireless communications and networking: A survey and perspectives. arXiv preprint, 2025.

[25] BARONA LÓPEZ L I, BORJA SALTOS T. Heterogeneity challenges of federated learning for future wireless communication networks. Journal of Sensor and Actuator Networks, 2025, 14(2): 37.

[26] JADAV N K, NAIR A R, GUPTA R, et al. AI-driven network softwarization scheme for efficient message exchange in IoT environment beyond 5G. International Journal of Communication Systems, 2025, 38(1): e5336.

[27] MOU D. A comprehensive survey of deep learning–based object tracking for augmented reality in complex real-world scenes. Applied Artificial Intelligence Research, 2026, 2(1).

[28] AOUEDI O, LE V A, PIAMRAT K, et al. Deep learning on network traffic prediction: Recent advances, analysis, and future directions. ACM Computing Surveys, 2025, 57(6): 1-37.

[29] YE R, XU Z, PANG J. DDFM: A novel perspective on urban travel demand forecasting based on the ensemble empirical mode decomposition and deep learning//Proceedings of the 5th International Conference on Big Data Technologies. New York: ACM, 2022: 373-379.

[30] YE S, WU Q, FAN P, et al. A survey on semantic communications in internet of vehicles. Entropy, 2025, 27(4): 445.

[31] SHA X, LI B, ZHAO Z, et al. How Does Population Aging Affect New Quality Productivity in Economic Sustainability? An Empirical Study Based on Mediating Mechanisms and Moderating Effects. Sustainability, 2025, 17(18): 8249.

[32] SUN H, LV Z, LI J, et al. Prediction of cancellation probability of online car-hailing orders based on multi-source heterogeneous data fusion//International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer Nature Switzerland, 2022: 168-180.

[33] LV Z, CHENG Z, LI J, et al. TreeCN: time series prediction with the tree convolutional network for traffic prediction. IEEE Transactions on Intelligent Transportation Systems, 2023, 25(5): 3751-3766.

[34] LI H, LI J, LV Z, et al. MFAGCN: multi-feature based attention graph convolutional network for traffic prediction//International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer International Publishing, 2021: 227-239.

[35] XU Z, LV Z, CHU B, et al. Fast autoregressive tensor decomposition for online real-time traffic flow prediction. Knowledge-Based Systems, 2023, 282: 111125.

[36] AMUGONGO L M, MASCHERONI P, BROOKS S, et al. Retrieval augmented generation for large language models in healthcare: A systematic review. PLOS Digital Health, 2025, 4(6): e0000877.