Enhancing Contextual Understanding in Knowledge Graphs: Integration of Quantum Natural Language Processing with Neo4j LLM Knowledge Graph

Suman Bharti
Computer Science Department
Kennesaw State University
Kennesaw, GA, USA
sbharti@students.kennesaw.edu

Dan Chia-Tien Lo
Computer Science Department
Kennesaw State University
Kennesaw, GA, USA
dlo2@kennesaw.edu

Yong Shi
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
yshi5@kennesaw.edu

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10826002

***Abstract*—Traditional Knowledge Graphs (KGs), such as Neo4j, face challenges in managing high-dimensional relationships and capturing semantic nuances due to their deterministic nature. Quantum Natural Language Processing (QNLP) introduces probabilistic reasoning into the KG context. This integration leverages quantum principles, such as superposition, which allows relationships to exist in multiple states simultaneously, and entanglement, where the state of one entity dynamically influences the state of another. This quantum-based probabilistic reasoning provides a richer, more flexible representation of connections, moving beyond binary relationships to model the nuances and variability of real-world interactions. Our research demonstrates that QNLP enhances Neo4j's ability to analyze context-rich data, improving tasks like entity extraction and knowledge inference. By modeling relationship states probabilistically, QNLP addresses limitations in traditional methods, providing nuanced insights and enabling more advanced, context-aware NLP applications.**

***Index Terms*—Neo4j, QNLP, LLM, KGs, NLP.**

I. Introduction

Neo4j's Large Language Model (LLM) Knowledge Graph is a graph-based database optimized for representing relationships derived from unstructured textual data, like resumes, documents or social media content. It excels in scalability, efficiently handling large datasets with millions of nodes and edges, and supports optimized querying through powerful query languages like Cypher for efficient data retrieval and analysis. Due to deterministic limitations, Neo4j lacks flexibility for handling context-dependent relationships effectively. Neo4j lacks probabilistic reasoning capabilities, which are essential for representing high-dimensional relationships involving multiple interconnected attributes and capturing semantic variability. Subtle differences in meaning or context, such as "works with" versus "collaborates closely with," are often lost in deterministic frameworks. This research addresses these shortcomings by integrating Quantum Natural Language Processing (QNLP) with Neo4j, enabling it to process and model relationships probabilistically for greater semantic depth and accuracy.

Quantum Natural Language Processing (QNLP) leverages quantum computing principles to address the limitations of deterministic models. Quantum concepts such as superposition and entanglement enable probabilistic reasoning, allowing relationships to exist in multiple states simultaneously, thereby offering a richer and more flexible representation of connections. For instance, a relationship can simultaneously reflect probabilities for both strong and weak collaboration. Entanglement ensures that changes in one entity's state dynamically influence the state of connected entities, capturing the complexity of context-rich relationships. By integrating QNLP, Neo4j LLM KG can process relationships probabilistically, moving beyond rigid deterministic models to achieve a nuanced understanding of complex, context-dependent relationships common in real-world data.

II. Literature Survey

Knowledge Graphs have been widely adopted for structuring and representing relationships between entities, as discussed by Ji et al. [2]. However, traditional NLP approaches and Neo4j LLM Graph Databases face computational and structural challenges in capturing the semantic depth of complex relationships. Quantum NLP, as proposed by Kartsaklis et al. [3], uses quantum mechanics to efficiently handle high-dimensional data and model intricate semantic relationships. Recent studies by Vijayakumar et al. [1] highlight the potential for QNLP in enhancing Knowledge Graph applications, particularly in overcoming limitations in traditional models.

A significant motivation for our current work is derived from the Compositional Distributional Model (DisCoCat) discussed in [5][6]. The DisCoCat model integrates categorical information and distributional semantics, allowing for the encoding of word and phrase meanings as quantum states. This model leverages the principles of quantum mechanics, such as superposition and entanglement, to represent and process linguistic information more effectively. By combining structural aspects of language (grammar theory and syntax) with statistical approaches based on empirical evidence, DisCoCat provides a robust framework for understanding and modeling natural language.

The Compositional Distributional Model has shown promise in addressing the limitations of current NLP models, which often require large datasets and extensive computational resources. By utilizing quantum circuits to represent linguistic structures, DisCoCat offers a more scalable and efficient approach to NLP tasks. This model's ability to handle complex, high-dimensional relationships and capture semantic nuances aligns well with the goals of our research, which aims to integrate QNLP with Knowledge Graphs to enhance contextual understanding.

In summary, the integration of QNLP with Knowledge Graphs builds on the foundational work of DisCoCat, leveraging quantum principles to improve the representation and processing of linguistic information. This approach addresses the limitations of traditional NLP methods and Neo4j LLM Graph Databases, providing a more nuanced and probabilistic understanding of entity relationships. Our research aims to further explore and validate the potential of QNLP in enhancing Knowledge Graph applications, ultimately leading to more advanced and context-aware NLP solutions.

III. Methodology

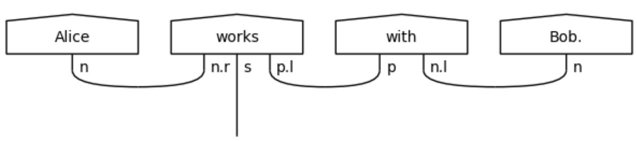

Fig. 1. DisCoCat representation of the sentence ”Alice works with Bob,” showing the grammatical structure where each word is connected based on its role. (nouns, verbs, and prepositions)

In this framework, relationship states are defined as probabilistic representations of connections between entities, moving beyond the fixed links used in deterministic models. Traditional representations, such as "Alice works with Bob," simply indicate the presence or absence of a relationship without accounting for the nuances or variability inherent in real-world interactions. In contrast, probabilistic states allow for multiple interpretations of a relationship, derived using quantum principles like superposition and entanglement. Superposition enables a relationship to exist in multiple states simultaneously, capturing varying degrees of collaboration or interaction. For example, the relationship "Alice works with Bob" could be represented as "00," signifying minimal or no collaboration, or "11," indicating strong or high collaboration. Entanglement further enhances this approach by ensuring that changes in one entity's state influence the state of related entities, reflecting the interconnected and dynamic nature of relationships.

This quantum approach generates a distribution of relationship states, offering deeper insights into varying levels of collaboration or connection. Such probabilistic representations capture the uncertainty and variability characteristic of real-world relationships, enriching the semantic depth of knowledge graphs like Neo4j. The integration of Quantum Natural Language Processing (QNLP) into Neo4j enables this probabilistic modeling through a structured pipeline. First, unstructured text, such as resumes, is processed into Neo4j's LLM Knowledge Graph. Then, relationships are parsed using lambeq's BobcatParser, converting sentences into DisCoCat diagrams that represent syntactic and semantic structures shown in Fig. 1. The DisCoCat diagram is then mapped to a quantum circuit using the IQPAnsatz, which defines atomic types (e.g., nouns, sentences) and maps them to qubits. The quantum circuit is created by converting the DisCoCat diagram into quantum gates, which are then simulated using Qiskit's Aer simulator. The quantum circuit is executed, and the results, representing the probabilistic outcomes of the relationships between entities, are obtained and visualized in a histogram. This process illustrates how QNLP can provide a richer, probabilistic interpretation of entity relationships, enhancing the contextual understanding within Knowledge Graphs. The final output, a probabilistic distribution of relationship states, offers deeper insights compared to traditional deterministic models.

Unlike classical models where a relationship is either present or absent, QNLP assigns probabilities to each potential state of the relationship, allowing for variability and context-dependent nuances. For instance, a traditional model might indicate that "Alice works with Bob" is either true or false, but QNLP provides a probabilistic distribution showing the likelihood of different collaboration levels. This advanced modeling reflects the complexity of real-world data, enabling Neo4j to represent relationships with greater depth, flexibility, and context-awareness.

IV. Preliminary Results and Analysis

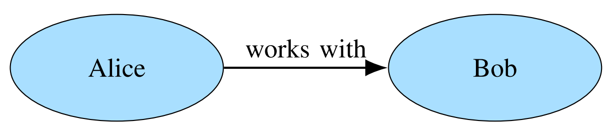

Fig. 2. A deterministic Neo4j representation of "Alice works with Bob."

In Neo4j LLM Knowledge Graph framework, Alice and Bob are represented as nodes, with a directed arrow labeled ”works with” indicating their relationship, as illustrated in Fig. 2. This straightforward representation captures the existence of a connection but fails to express deeper nuances, such as the strength, variability, or probabilistic nature of their collaboration. Such deterministic models lack the ability to account for real-world uncertainties and context-dependent relationships. This limitation is addressed by Quantum Natural Language Processing (QNLP), as illustrated in Fig. 3, which shows the output of quantum simulations for the relationship "Alice works with Bob." The simulation results indicate two primary probabilistic states: "00", representing no collaboration, and "11", indicating active collaboration. The near-equal probability distribution of these states demonstrates the probabilistic nature of QNLP, which captures the inherent variability and uncertainty of real-world relationships.

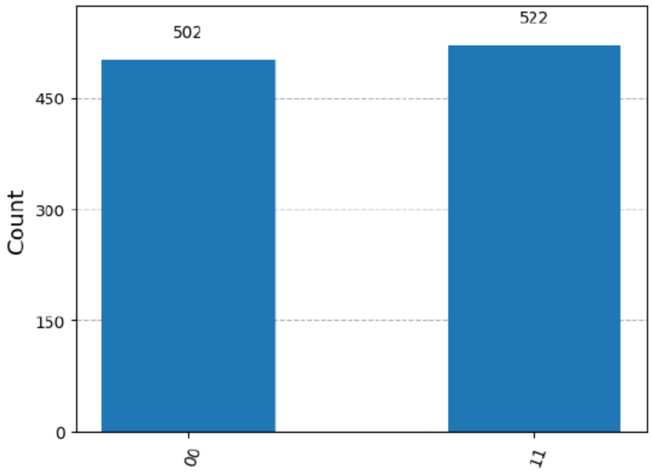

Fig. 3. The quantum simulations generated probabilistic outcomes for relationships, such as “Alice works with Bob.” The results showed two main states: “00” and “11”. Here, “00” represents a scenario where no collaboration exists between Alice and Bob, while “11” indicates active collaboration. The near-equal probability of these states highlights the probabilistic nature of QNLP, capturing the inherent variability in real-world relationships.

Traditional knowledge graphs (KGs) struggle to manage uncertainty and ambiguity in relationships, relying instead on deterministic links that are either present or absent. By contrast, QNLP introduces probabilistic reasoning, enabling Neo4j to capture nuances and context-rich relationships more effectively. Probabilistic states allow the relationship between Alice and Bob to be expressed as a range of possibilities, rather than a binary connection, thereby offering greater flexibility in representing real-world complexities. For example, instead of simply indicating that “Alice works with Bob,” QNLP provides a distribution of probabilities reflecting varying levels of collaboration strength. This enriches Neo4j’s semantic reasoning capabilities, enabling it to distinguish between strong and weak relationships, handle ambiguous connections, and better analyze entity interactions.

The comparative advantage of QNLP lies in its ability to enhance Neo4j’s functionality by capturing these probabilistic insights. For instance, while traditional KGs may fail to account for overlapping or context-dependent relationships, QNLP models these with a richer framework that reflects the subtleties of real-world data. This integration transforms Neo4j into a more advanced tool for analyzing relationships, improving tasks like entity extraction, knowledge inference, and semantic reasoning. By providing probabilistic measures of relationship strength, QNLP ensures that Neo4j can represent complex, dynamic, and context-aware relationships more accurately than classical methods.

V. Future Work and Directions

The integration of Quantum Natural Language Processing (QNLP) with Knowledge Graphs (KGs) opens up numerous possibilities for future research and applications across various domains. In e-commerce, QNLP can enhance personalized recommendation systems by analyzing complex relationships between users and products, considering contextual nuances and probabilistic patterns in user behavior. In workforce planning, QNLP can be used to identify skill gaps, model dynamic team interactions, and provide insights into improving team performance by analyzing probabilistic relationships between employees, their roles, and organizational goals. These applications demonstrate the potential of QNLP to revolutionize industries by offering deeper, context-aware insights.

VI. Conclusion

The integration of Quantum Natural Language Processing (QNLP) with Neo4j LLM Knowledge Graphs (KGs) marks a significant leap forward in relationship modeling for natural language processing tasks. By introducing probabilistic reasoning, QNLP overcomes the deterministic limitations of Neo4j, enabling it to manage high-dimensional relationships and capture previously unattainable semantic nuances. Leveraging quantum principles such as superposition and entanglement, QNLP allows Neo4j to process relationships probabilistically, providing a richer, context-aware representation of connections. This enhancement improves Neo4j’s performance in critical areas, including modeling contextual and overlapping relationships, extracting entities with greater precision by differentiating strong and weak links, and enhancing knowledge inference through nuanced predictions and insights. These advancements enable Neo4j to handle ambiguous, multi-faceted relationships more effectively, making it a more robust tool for advanced NLP applications. By bridging the gap between quantum computing and knowledge graphs, this research opens the door to innovative solutions across domains such as healthcare, e-commerce, and workforce planning, where understanding complex relationships is critical.

Acknowledgement

The work is partially supported by the U.S. National Science Foundation under awards #2413540 and #2100115. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Vijayakumar, S., Louis, F. (2023). Revolutionizing Staffing and Recruiting with Contextual Knowledge Graphs and QNLP: An End-to-End Quantum Training Paradigm. Proceedings of the IEEE International Conference On Knowledge Graph (ICKG).

[2] Ji S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S. (2021). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. IEEE Trans. Neural Netw. Learn. Syst.

[3] Kartsaklis, D., Fan, I., Yeung, R., et al. (2021). Lambeq: An Efficient High-Level Python Library for Quantum NLP. ArXiv.

[4] Preskill, J. (2018). Quantum Computing in the NISQ Era and Beyond. Quantum.

[5] Guarasci, R., De Pietro, G., Esposito, M. (2022). Quantum natural language processing: Challenges and opportunities. Applied sciences, 12(11), 5651.

[6] Coecke, B., Sadrzadeh, M., Clark, S. (2010). Mathematical foundations for a compositional distributional model of meaning. arXiv preprint arXiv:1003.4394.