Peer Review

Review of: "Top Ten Challenges Towards Agentic Neural Graph Databases"

Huy Truong[1]

1. University of Groningen, Netherlands

Summary

The authors discuss Agentic Neural Graph Databases (ANGDBs), which integrate advanced functionalities into Neural Graph Databases (NGDBs)—graph databases enhanced by Graph Neural Networks (GNNs). The manuscript outlines ten key challenges in implementing ANGDBs, including: (1) Semantic Units, (2) Abductive Reasoning, (3) Query Families Generalization, (4) Privacy and Security, (5) Scaling Laws, (6) Distributed System Integration, (7) Compatibility with Traditional Graph Databases, (8) Natural Language Grounding, (9) Adapting NGDBs to Large Language Models (LLMs), and (10) Smart Applications.

ANGDB extends NGDB into an autonomous system with advanced capabilities in reasoning, self-learning, and scalability. Challenges (1)–(3) focus on enhancing inference, reasoning, and generalization. Challenges (4)–(6) address privacy, scalability, and distributed system concerns—key requirements for modern Software as a Service (SaaS) solutions. Furthermore, integrating LLMs (Challenge 9) strengthens ANGDB's reasoning ability, surpassing traditional graph databases (Challenge 8). Finally, Challenge 10 highlights potential applications in data management, recommendation systems, and event processing. Ultimately, ANGDB aims to function as an autonomous agent, directly managing databases while interacting with users and external systems.

Strengths:

The authors highlighted ANGDBs as an example of the "next-gen" database that not only handles genuine data but also devises new information (embeddings, hidden relationships, etc.) organized in a graph structure. Also, they acknowledged the concern of privacy and security and, therefore, proposed

 doi.org/10.32388/H26IK0

techniques to prevent malicious attacks as well as strengthen the defense at the neural model level. Similarly, as addressed in Challenge 7, it is essential to have a standard benchmark to compare ANGDBs with other traditional databases or ANGDBs' versions improved from different sources of knowledge.

Weaknesses:

There are a few disagreements, including (1) the definition of an agent and (2) complete autonomy. As in the manuscript, ANGDB is mentioned as NGDBs extended with additional functionalities such as data permutation and management. However, NGDB is also involved in such functions (constructing labels given node and edge context— a form of data permutation). Here, the term 'Agent' is not fully defined and could benefit from further clarification to address basic questions such as: What constitutes an agent? How many agents exist in the database? Do they operate independently or cooperatively?

Another concern is ANGDB's ability to CREATE, UPDATE, and DELETE data in the database. Since the manuscript does not specify which types of data are manipulated, there is a risk that essential primary data, such as nodal and edge features, could be unexpectedly removed due to the inherent unpredictability of probabilistic models. Additionally, generating new data, queries, or knowledge may negatively impact the system's sustainability and efficiency. This issue should be recognized and addressed as a significant challenge.

In terms of writing, the manuscript is well-written and easy to follow. However, abbreviations (e.g., GNNs, NGDB) should be introduced only once at the first mention and not repeated unnecessarily.

Suggestion

It is worth mentioning the risk of fusing grounding knowledge out of the scope of the existing database, as this could drive the neuron models to a different end. Also, the management ability of ANGDBs between knowledge devised from itself and multiple versions should be considered a challenge because of the expansion of the computational and storage units. Last but not least, despite the autonomy priority, it is essential to include mechanisms for human intervention to handle unintended outcomes.

Declarations

Potential competing interests: No potential competing interests to declare.