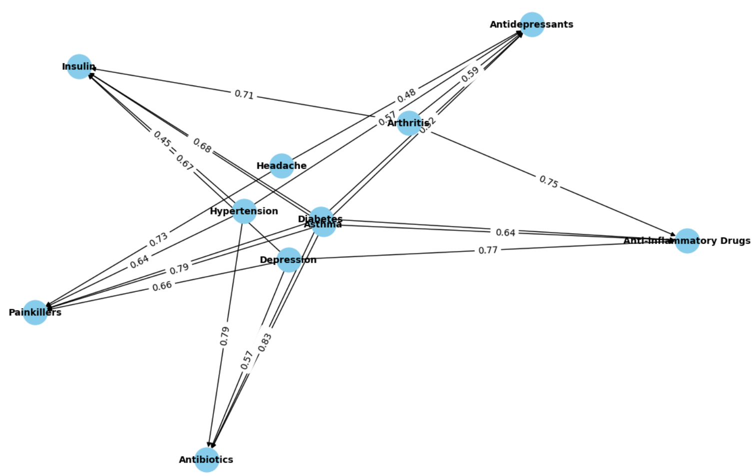
![PHOTREK]

Architectural Design for Uncertain Knowledge Graphs

Milestone 4 Report



Principal Investigator & Project Manager:

Kenric Nelson

Photrek, LLC

admin@photrek.io

12 April, 2024

Executive Summary

Photrek uses Neo4j as its graph database management system. Neo4j is an industry leader and defacto standard for graph databases and using Cypher for graphical element creation allows for extensions to uncertain knowledge graph approaches. It has been successfully employed on similar projects. Python and networkx for its extensive libraries and tools that facilitate data processing and visualization.

Our use case builds and studies the relationship between several diseases and their treatments via medication. Probabilistic reasoning is introduced with emphasis on risk aware techniques that have been under study by Photrek researchers for several years.

Amenah AL-Najafi provided a medical dataset from a hospital in the Middle East consisting of disease-treatment pairs and deduced the probability of each treatment's suitability for a given disease. Diseases and treatments are represented as nodes, and the probabilities were captured as edges between them in the graph. We imported the collected medical data into the Neo4j database, creating nodes for diseases and treatments and establishing relationships between them based on the collected data. We employed Python libraries to visualize and analyze the data, illustrating the relationships between diseases and medications, as well as the effectiveness of treatments across various diseases.

Milestone 4 Budget & Objectives

Total Budget: $ 6,550 USD in AGIX
Objectives:
- Define code base foundation
- Choose specifics of use case
- Complete analysis of architecture with use case

Deliverable: Report on Use Case with bounds on uncertainty

1. Define code base foundation

In establishing the foundation of our database, Photrek will use Neo4j as a graph database management system. Neo4j is a reliable, highly efficient graph database that excels in handling interconnected data and has been used successfully in similar applications (Navarrete and Vallecillo 2021). This makes it particularly suitable for storing the medical data in our use case and efficiently querying it via Cypher to extract potential relationships in a manner that reflects real-world connections and implications. We will use Python and its associated libraries such as networkx and others for its extensive libraries and tools that facilitate data processing and visualization.

2. Choose specifics of use case

Photrek's use case will use acquired medical data to elucidate the relationship between diseases, their treatments, and treatment effectiveness. Emphasis was placed on acquiring comprehensive data regarding diseases and their treatments, while also assessing the likelihood of each treatment's suitability for specific diseases. This is an area of great impact and current interest (Liu et al. 2024) (Chandak, Huang, and Zitnik 2023).

During the study, significant attention was dedicated to computing coupled probabilities, recognized as a fundamental aspect of risk-aware machine intelligence. A methodology was employed to specifically calculate coupled probabilities in a scenario with a degree of freedom of approximately 999 (sample size -1), The successful execution of the methodology highlighted the capability to compute coupled probabilities effectively, even in scenarios with substantial degrees of freedom.

Through the incorporation of computed relative risk aversion (r) into the calculations, the coupled probabilities provided a cautious assessment of uncertainty.

Methodology:

1. The calculation of coupled probabilities involves determining several parameters, including the relative risk aversion (r), which plays a crucial role in reshaping the distribution based on uncertainty.
   In this analysis, the parameters are set as follows:
   - $\alpha = 2$ (for two-sided distributions).
   - d = 1 (dimensions of the distribution).
   - $\lambda = 999$ (degree of freedom).
   - κ (inverse of the degree of freedom).
2. Using the formula $r = \frac{\alpha\kappa}{1+d\kappa}$, the relative risk aversion (r) (Nelson 2020) is computed based on the provided parameters.

3. Once the relative risk aversion (r) is determined, it is used in the computation of coupled probabilities by raising the probabilities to the power of the relative risk aversion (r) and then renormalizing them.

3. Complete analysis of architecture with use case

- We collected medical data consisting of disease-treatment pairs along with the probability of each treatment's suitability for a given disease.
- Using Neo4j, we designed a graph database schema to represent the relationships between diseases and treatments. Each disease and treatment was represented as a node, and the probabilities were captured as edges between them in the graph.
- We imported the collected medical data into the Neo4j database, creating nodes for diseases and treatments and establishing relationships between them based on the collected data.
- We employed Python libraries to visualize and analyze the data, illustrating the relationships between diseases and medications, as well as the effectiveness of treatments across various diseases.

4. Joining New England Research Cloud

The Photrek team has established an account with the New England Research Cloud (NERC). Raymond Mata has joined Photrek as a Systems Administrator. Ray brings experience as the founder and operator of the Cardano GROW Pool and is currently supporting the SingularityNET Hypercycle team in its roll out of computational infrastructure for decentralized AI. Ray's initial work is supporting the Simulating Risky World project to host Photrek's AI Marketplace algorithms. Once this is complete he will work with the Uncertain Knowledge Graph team to support their computational requirements.

Project Plans

Budget & Schedule

Note: The milestone numbers were revised per SingularityNET's request to start at 1.

M#	Milestone name	Milestone deliverable	Related budget	Status
1	**Contract Signing & Management Reserve**	**Funding put in reserve to address unforeseen requirements** **Contract Signature**	**$ 1,750.00**	**Report and Invoice Submitted on Dec 7th.**
2	**Brief SingularityNET KG & LLM team regarding Objectives**	• **Determine current status of SingularityNETs KG & LLM development** • **Acquire feedback on probability with upper and lower bounds** • **Acquire feedback on use of UKG as augmentation to static KG & LLM** **Report on refined objectives**	**$ 5,460.00**	**Report & Invoice submitted Dec 22nd.**
3	**Draft Architecture Design**	• **Design details of use of probabilistic graph to form UKG** • **Design details of use** • **Draft architectural plan** **Report with architecture design**	**$ 5,870.00**	**Feb 16th**
4	**Develop Use Case Plan**	• **Define code base foundation** • **Choose specifics of use case** • **Complete analysis of architecture with use case** **Report on opportunity risk analysis**	**$ 6,550.00**	**Plan: Apr 12th**
5	Develop UKG	• Use probabilistic graph foundation to create UKG facts • Develop small network of UKG facts • Develop interface with static KG and/or LLM Report with initial code	$ 6,550.00	Plan: Apr 26th
6	Use Case Demo	• Demo using UKG independently • Demon using upper & lower probabilities • Demo using UKG with static KG and/or LLM Report on Use Case Demo	$ 7,550.00	Plan: Jun 19th
7	Final Report with Next Steps	• Document strengths and weaknesses of design • Document lessons learned from use case • Document plans for next steps	$ 6,270.00	Plan: Jul 24th

		Final Report with next steps		

References

Chandak, Payal, Kexin Huang, and Marinka Zitnik. 2023. "Building a Knowledge Graph to Enable Precision Medicine." *Scientific Data* 10 (1): 67. https://doi.org/10.1038/s41597-023-01960-3.

Liu, Qi, Qinghua Zhang, Fan Zhao, and Guoyin Wang. 2024. "Uncertain Knowledge Graph Embedding: An Effective Method Combining Multi-Relation and Multi-Path." *Frontiers of Computer Science* 18 (3): 183311. https://doi.org/10.1007/s11704-023-2427-z.

Navarrete, Francisco J., and Antonio Vallecillo. 2021. "Introducing Subjective Knowledge Graphs." In *2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC)*, 61–70. Gold Coast, Australia: IEEE. https://doi.org/10.1109/EDOC52215.2021.00017.

Nelson, Kenric P. 2020. "Reduced Perplexity: A Simplified Perspective on Assessing Probabilistic Forecasts." In *Advances in Info-Metrics: Information and Information Processing across Disciplines*, edited by Min Chen, J. Michael Dunn, Amos Golan, and Aman Ullah, 0. Oxford University Press. https://doi.org/10.1093/oso/9780190636685.003.0012.