Deep dive into Machine learning by leveraging Networks

WE Machine Learning Project

Aarushi Gulati, Samhitha Bharthulwar, Saumya Chaturvedi

WE Program Cohort 4

December 20, 2023

Project Idea

A deep dive into knowledge graphs and GNNs.

- We've done basic ML
- Graphs/Networks are very powerful - recommeder systems, PageRank, AlphaFold, drug side effects, traffic prediction, weather forecasting
- Modelling relation data
- Generate powerful, useful visual representations

Getting familiar with some terms

Heterogenous graph: A graph with multiple types of nodes and/or relations.

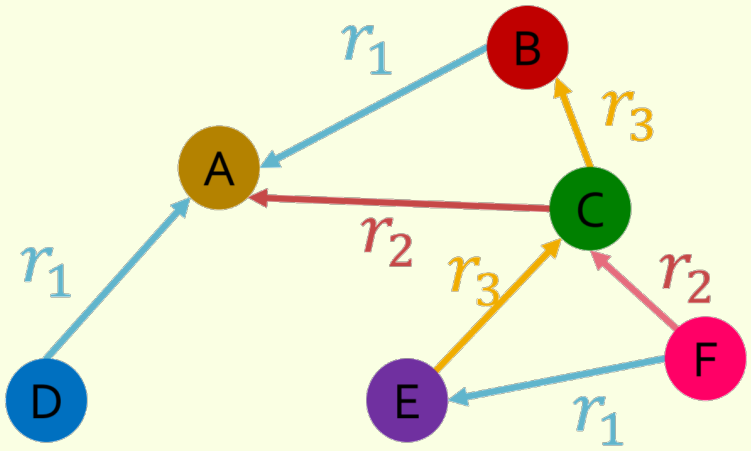

Figure 1: An example of a heterogenous graph [2]

Getting familiar with some terms

Knowledge graph: A heterogenous graph where nodes represent entities and are labelled with types, and edges capture the relation between entities.

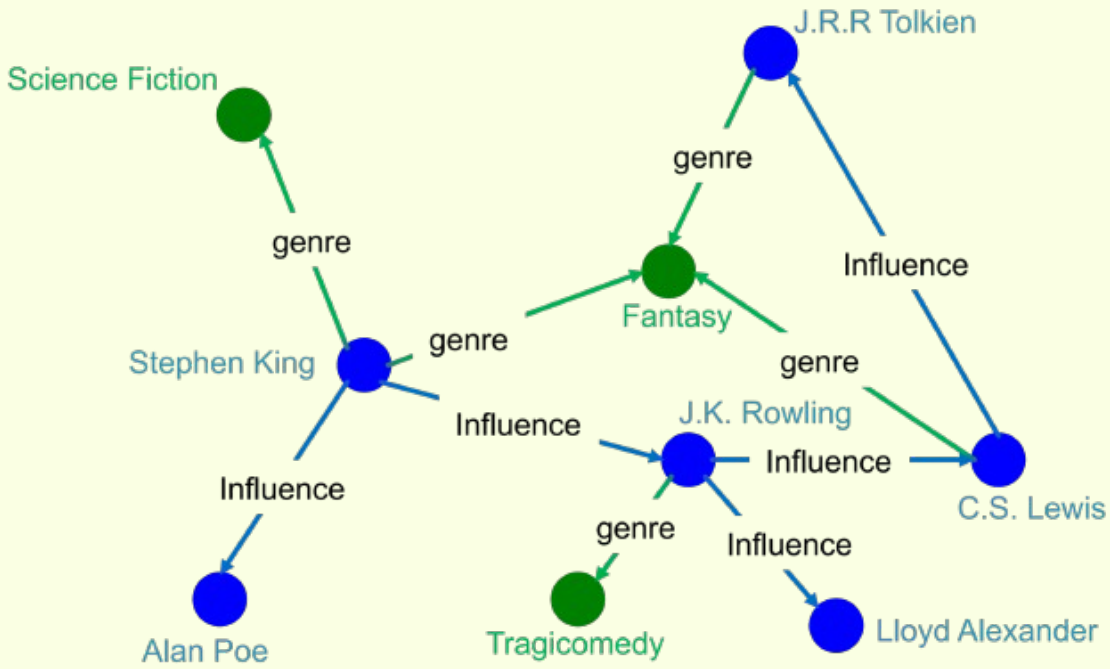

Figure 2: An example of a knowledge graph [2]

Current Progress

- Studying networks [2]
- Key steps in building a knowledge graph - Named Entity Recognistion (NER) + Relation Classification (RC)
- Experimenting - REBEL, Mistral 7B, Llama

Learning Objectives

- Represent data corpus through knowledge graphs [1]
- Explore Graph Visualisation techniques
- Work with Knowledge Graph Transformers
- Learn about networks and their properties
- Apply different graph neural networks

Strategy

- Input: complex knowledge base
- Generate a knowledge graph: NER + RC
- Visualisation
- Perform Link Prediction, Entity Classification

Next Steps

- Continue exploring different ways to create KGs.
- Try different Graph Visualization methods.
- Implement KG embeddings (using BERT KG and other Transformers).
- Learning about the working of GNNs.

Timeline

- Bare minimum: KG creation + visualization + 1 GNN application
- Satisfactory: if we complete KG creation, visualization and application of 2-3 GNNs
- Excelling: various applications with code and proper documentation for all 3 stages, while accomplishing 2-3 tasks with the dataset

References

[1] Linmei Hu et al. "Text-Graph Enhanced Knowledge Graph Representation Learning". In: *Frontiers in Artificial Intelligence* 4 (2021). ISSN: 2624-8212. DOI: 10.3389/frai.2021.697856. URL: https://www.frontiersin.org/articles/10.3389/frai.2021.697856.

[2] Stanford University Jure Leskovec. *CS224W: Machine Learning with Graphs*. 2021. URL: https://snap.stanford.edu/class/cs224w-2021/.