Homework 7: KG Embedding
DSCI 558 BUILDING KNOWLEDGE GRAPH

DUE DATE: Sunday, 03/28/2021 @ 11:59pm on Blackboard

Ground Rules

This homework must be done individually. You can ask others for help with the tools, but the submitted homework needs to be your own work.

Summary

In this homework, you will learn about KG embedding and how to use AmpliGraph to generate KG embeddings using different models. You will need to run your homework on Python notebooks. Please make sure you have output for all notebook code cells when converting to pdf.

Task 1 – AmpliGraph tutorial (4pts)

Install and complete the AmpliGraph's "Basic Tutorial" (https://docs.ampligraph.org/en/1.3.2/tutorials/AmpliGraphBasicsTutorial.html) on a Python notebook. It is recommended to use Python 3.7 as shown in the tutorial.

Train other KG embedding methods: TransE and DistMult. Compare the result of three methods in your notebook using: MRR score, hit@10, hit@3 and hit@1.

Task 2 – KG embedding model (6 pts)

Download the remapped fb15k_237 dataset here: https://ampgraphenc.s3-eu-west-1.amazonaws.com/datasets/freebase-237-merged-and-remapped.csv

Split the data to create a test set of 10000 triples. Create another Python notebook and train a KG embedding model of your choice (should be the best performing method from the tutorial) and report its performance on the test set using: MRR score, hit@10, hit@3 and hit@1. Show your model's performance in the notebook.

Randomly select 10 different triples to run the prediction and show the predicted result in your notebook using the format below.

	statement	rank	score	prob
5	Daenerys Targaryen SPOUSE Craster	4090	-2.750880	0.060037
10	Brandon Stark ALLIED_WITH House Lannister of C...	3821	-1.515378	0.180143
18	Cersei Lannister PARENT_OF Brandon Stark	4061	-1.386417	0.199980
1	Tyrion Lannister SPOUSE Missandei	3190	-0.554477	0.364826

Hint: If your training takes too long to complete (you can estimate by running time per epoch), you can use the following solutions and explain in your notebook:

- Reduce the number of epochs
- Use Google Colab for your training
- Split your data using train-test-split-no-unseen (remember to save so that you can reuse if necessary)

Submission Instructions

You must submit the following files a single .zip archive named Firstname_Lastname_hw6.zip and submit it via Blackboard:

- Firstname_Lastname_hw7_task_{task_number}.ipynb: A Python notebook containing answers for Task 1 and Task 2
- Firstname_Lastname_hw7_task_{task_number}.pdf: A printed pdf version of the notebook for Task 1 and Task 2