Assignment 2: Link Prediction with NASA GES-DISC Dataset

Objective: Apply two link prediction methods to the NASA GES-DISC knowledge graph dataset.

Part 1: Exploring the NASA GES-DISC Dataset (20%)

- **Task:** Download and explore the NASA GES-DISC dataset.
- **Dataset Link:** https://zenodo.org/record/11492533
- **Description:** Analyze node types, edges, and relationships. Provide basic statistics, including node and edge counts and any significant relationships.

Deliverable:

- Code to load the dataset using PyTorch Geometric (PyG) or another library.
- Print summary of the dataset's structure and key statistics.

Part 2: Link Prediction (60%)

1. **Method 1: Embedding-Based Approach (30%)**
    - **Task:** Apply an embedding-based method for link prediction.
    - **Description:** Train a model that generates node embeddings, then use those embeddings to predict links. Print relevant metrics.
2. **Method 2: Alternative Approach (30%)**
    - **Task:** Choose and implement another link prediction method.
    - **Description:** This method should not use embeddings. You can use any approach of your choice. Compare the performance of this method with the embedding-based method.

Deliverable:

- Code for both methods.
- Print performance metrics for both approaches.
- Provide a comparison in terms of accuracy and computational efficiency.

Assignment 2: Link Prediction with NASA GES-DISC Dataset

Part 3: Reflection and Analysis (20%)

- **Task:** Compare both methods.
- **Description:** Write a reflection on the performance of each method. Discuss any challenges and insights gained. Suggest improvements to the dataset or the methods used.

Deliverable:

- A markdown cell(s) with your analysis and reflections in your jupyter notebook.

Grading Breakdown:

- **Dataset Exploration (Part 1):** 20%
- **Embedding-Based Link Prediction (Part 2 - Method 1):** 30%
- **Alternative Link Prediction (Part 2 - Method 2):** 30%
- **Reflection (Part 3):** 20%

Submission:

- Submit the full Jupyter notebook with all cells executed and results included.