**Model Performance Report: Drug Recommendation System**

**1. Introduction**

Drug repurposing is an important task in biomedical research, seeking to find drugs that are already available to treat various diseases. This project utilizes Knowledge Graph Embeddings (KGE) and Graph Neural Networks (GNNs) to produce alternative drug recommendations with safety constraints (side effects). Our method uses the Hetionet dataset and measures performance based on HITS@3 and NDCG@3.

**2. Methodology**

**2.1 Dataset**

We utilized the Hetionet knowledge graph, which has connections among drugs, diseases, and other biomedical entities. The data was divided into training, validation, and test sets.

**2.2 Model Architecture**

* Knowledge Graph Embeddings: We employed a TransE model with PyKEEN to learn drug-disease relationships.
* Drug Similarity Computation: We calculated cosine similarity between learned representations to rank substitute drugs.
* Filtering for Side Effects: Drugs with more than a set side effect threshold were removed.

**2.3 Evaluation Metrics**

1. HITS@3: Tests if at least one of the top 3 predicted drugs is found in the ground truth.
2. NDCG@3: Measures ranking quality based on the position of relevant drugs within the top 3 recommendations.

**3. Results & Analysis**

**3.1 HITS@3 Performance**

After removing drugs with too many side effects, we got:



This shows that our model was able to correctly identify at least one correct alternative drug per disease.

**3.2 NDCG@3 Performance**

NDCG@3 values were calculated to measure ranking quality and demonstrated strong agreement with ground truth rankings.

**4. Limitations & Future Work**

**4.1 Limitations**

* Dataset Constraints: The Hetionet dataset is not guaranteed to contain all drug-disease interactions.
* Side Effect Filtering: The cut-off for filtering out drugs is manually defined and might be optimized.
* Embedding Limitations: TransE is less capable of handling complex relationships than other models such as ConvE or RotatE.

**4.2 Future Improvements**

* Using state-of-the-art GNNs such as GraphSAGE to more effectively capture the graph structure.
* Investigating multi-modal learning by combining clinical trial information and gene expression profiles.
* Increasing interpretability through rationale for recommendations.

**5. Conclusion**

This project illustrates the efficacy of Knowledge Graph Embeddings for drug repurposing. With side effect constraints, we enhanced drug safety without compromising high recommendation accuracy (HITS@3 = 1.0000). Refining filtering mechanisms and investigating more sophisticated graph-based models will be the focus of future work.

**6. References**

1. Hetionet: A Biomedical Knowledge Graph([Hetionet - An integrative network of biomedical knowledge](https://het.io/))
2. PyKEEN: Knowledge Graph Embedding Framework([PyKEEN — pykeen 1.11.0 documentation](https://pykeen.readthedocs.io/en/stable/))
3. Graph Neural Networks in Drug Discovery([2003.04702](https://arxiv.org/pdf/2003.04702))