

Abstract

Link Prediction (LP) in Knowledge Graphs (KGs) is a fundamental task in machine learning and artificial intelligence, focused on inferring missing links by learning latent representations of entities and relations. This thesis presents a comprehensive comparative study of six prominent LP models TransE, ComplEx-N3, TuckER, Simple, TorusE, and CrossE evaluated across four widely-used benchmark datasets: FB15K, FB15K-237, WN18, and WN18RR.

These models span a range of architectural paradigms, including translational approaches (TransE, TorusE), bilinear and tensor factorization techniques (ComplEx-N3, TuckER), symmetric embeddings (Simple), and interaction-based mechanisms (CrossE). Each model is assessed using standard metrics: Mean Reciprocal Rank (MRR) and Hits@K ($K = 1, 3, 10$), with emphasis on both predictive performance and sensitivity to dataset properties such as symmetry, inverse relations, and structural redundancy.

The results reaffirm TransE's value as a computationally efficient baseline well suited for hierarchical and one-to-one relations, but less capable in modeling complex patterns like symmetry or many-to-many mappings. TuckER exhibits superior generalization, particularly on filtered datasets like FB15K-237 and WN18RR, while ComplEx-N3 excels on datasets rich in inverse or symmetric relations (e.g., FB15K, WN18), aided by complex-valued embeddings and N3 regularization. Simple underperforms in filtered, asymmetric contexts. CrossE delivers stable, moderate results across all datasets, reflecting its adaptability in noisy or sparse graphs. TorusE, while efficient in design, struggles with deep semantic interactions due to geometric limitations.

Overall, this study highlights the importance of aligning model architecture with dataset characteristics. It offers actionable insights for model selection and deployment in knowledge graph applications and contributes to a clearer understanding of the practical trade-offs in KGE research.