

# Reproducibility Report for ACM SIGMOD 2022 Paper: “Explaining Link Prediction Systems based on Knowledge Graph Embeddings”

HARIDIMOS KONDYLAKIS, FORTH-ICS, Greece

The experiments are reproducible and support the key findings of the paper. All experiments provide performance numbers similar to those from the paper.

## 1 INTRODUCTION

In this report, we describe the reproducibility effort for the SIGMOD 2022 paper “Explaining Link Prediction Systems based on Knowledge Graph Embeddings” [2] which is a joint work of Andrea Rossi, Donatella Firmani, Paolo Merialdo and Tommaso Teofili.

This work introduces Kelpie for explainable link prediction based on Knowledge Graph embeddings. Kelpie can be applied to any embedding-based link prediction models independently from their architecture, and it explains predictions by identifying the combinations of training facts that have enabled them.

## 2 SUBMISSION

In the paper, a link to a GitHub repository [1] is provided with a reproducibility package that allows the replication of all experiments included in the paper.

The reproducibility repository contains scripts for setting up the experimental environment, cloning and building the source code, downloading and preparing the datasets, running the experiments, and regenerating all the plots. The repository provides detailed instructions on how to run these scripts in either fully-automated or step-by-step mode.

## 3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1 shows the hardware and OS environments reported in the paper, and used in the reproducibility evaluation. The difference in the hardware specification affects raw performance numbers but not the overall trends and key findings.

Table 1. Hardware & Software environment

	Paper	Repro Review
CPU	88 CPUs Intel Core(TM) i7-3820 at 3.60GHz	16 CPUs AMD EPYC 7251 at 2.1 GHz
RAM	516GB	64GB
GPU	4 NVIDIA Tesla with 16GB VRAM	4 NVIDIA 2080ti with 11GB VRAM
OS	Ubuntu 18.04	Ubuntu 18.04
CUDA version	11.2	9.1

## 4 REPRODUCIBILITY EVALUATION

### 4.1 Process

We followed the step-by-step instructions as provided in the reproducibility repository. Repeating the whole body of experiments reported in the paper was discouraged by the authors as this was extremely time-consuming - even just limiting to the end-to-end experiments would exceed 4 months of uninterrupted run. Instead, a faster selection of representative experiments was

reproduced as suggested by the authors. A minor challenge is that required libraries were not installed at the beginning of the process. The problem arose only after a few hours, requiring a rerun of the whole script.

## 4.2 Results

All major findings of the paper in the representative experiments could be reproduced. More specifically the following end-to-end experiments and minimality experiments on the following combinations of models, datasets and scenarios were run:

- (1) ComplEx model, WN18 dataset, necessary scenario
- (2) ComplEx model, WN18 dataset, sufficient scenario
- (3) ComplEx model, FB15k-237 dataset, necessary scenario
- (4) ComplEx model, FB15k-237 dataset, sufficient scenario
- (5) ConvE model, WN18RR dataset, necessary scenario
- (6) ConvE model, WN18RR dataset, sufficient scenario
- (7) TransE model, FB15k dataset, necessary scenario
- (8) TransE model, FB15k dataset, sufficient scenario
- (9) TransE model, YAGO3-10<sup>6</sup> dataset, necessary scenario
- (10) TransE model, YAGO3-10<sup>6</sup> dataset, sufficient scenario

## 5 SUMMARY

Overall most paper claims have been reproduced with minimal effort. This is a stellar entry in SIGMOD reproducibility, setting a high standard on experiment-heavy papers at SIGMOD.

## REFERENCES

- [1] [n. d.]. Kelpie. <http://https://github.com/AndRossi/Kelpie>. Accessed: 2023-01-28.
- [2] Andrea Rossi, Donatella Firmani, Paolo Merialdo, and Tommaso Teofili. 2022. Explaining Link Prediction Systems based on Knowledge Graph Embeddings. In *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, Zachary Ives, Angela Bonifati, and Amr El Abbadi (Eds.). ACM, 2062–2075. <https://doi.org/10.1145/3514221.3517887>