

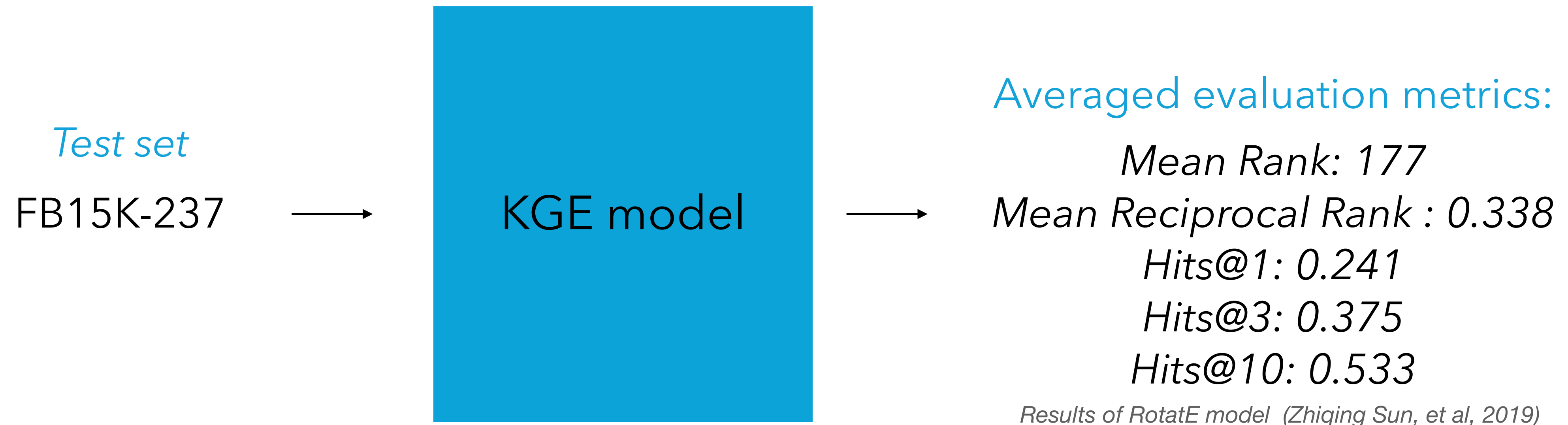
Behavioral Testing of Knowledge Graph Embedding Models

*Wiem Ben Rim, Carolin Lawrence, Kiril Gashteovski,
Mathias Niepert, Naoaki Okazaki*



Link Prediction Evaluation

Returns averaged evaluation metrics calculated over a given test set.



Typically, if $MRR_{\text{Model A}} > MRR_{\text{Model B}}$
then Model **A** is better than **B**

Motivation

Link Prediction Evaluation

What is the model **good** at?

What is the model **bad** at?

Motivation

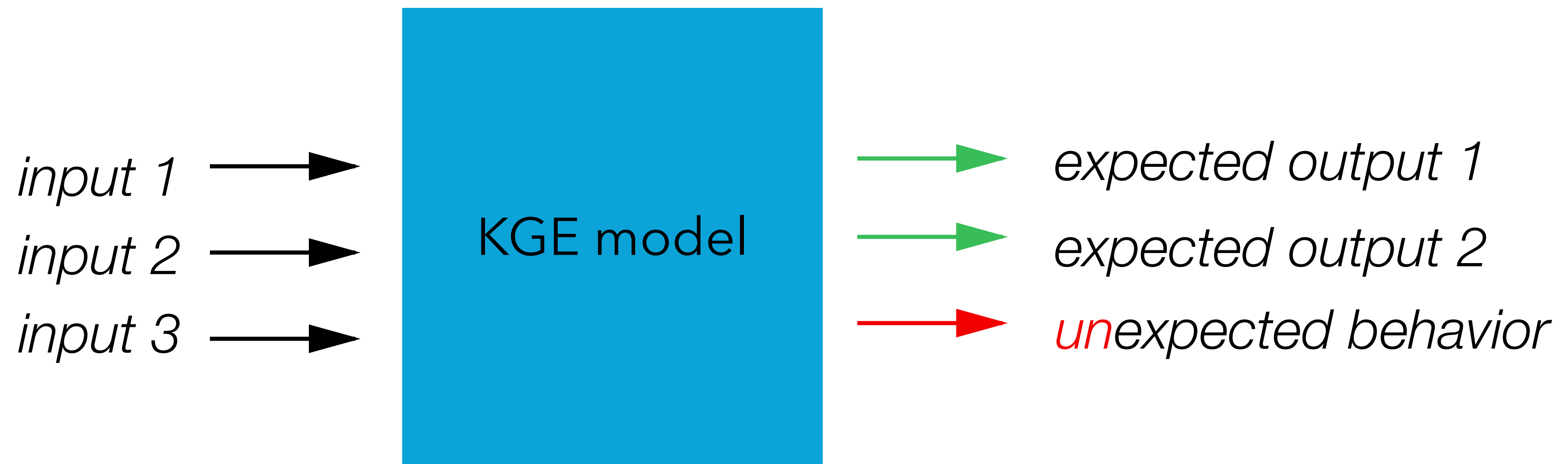
Behavioral Testing

Tests different capabilities of a system **without any knowledge of the internal structure** (Beizer, 1995).



Proposed Method

Behavioral Testing for KGE models



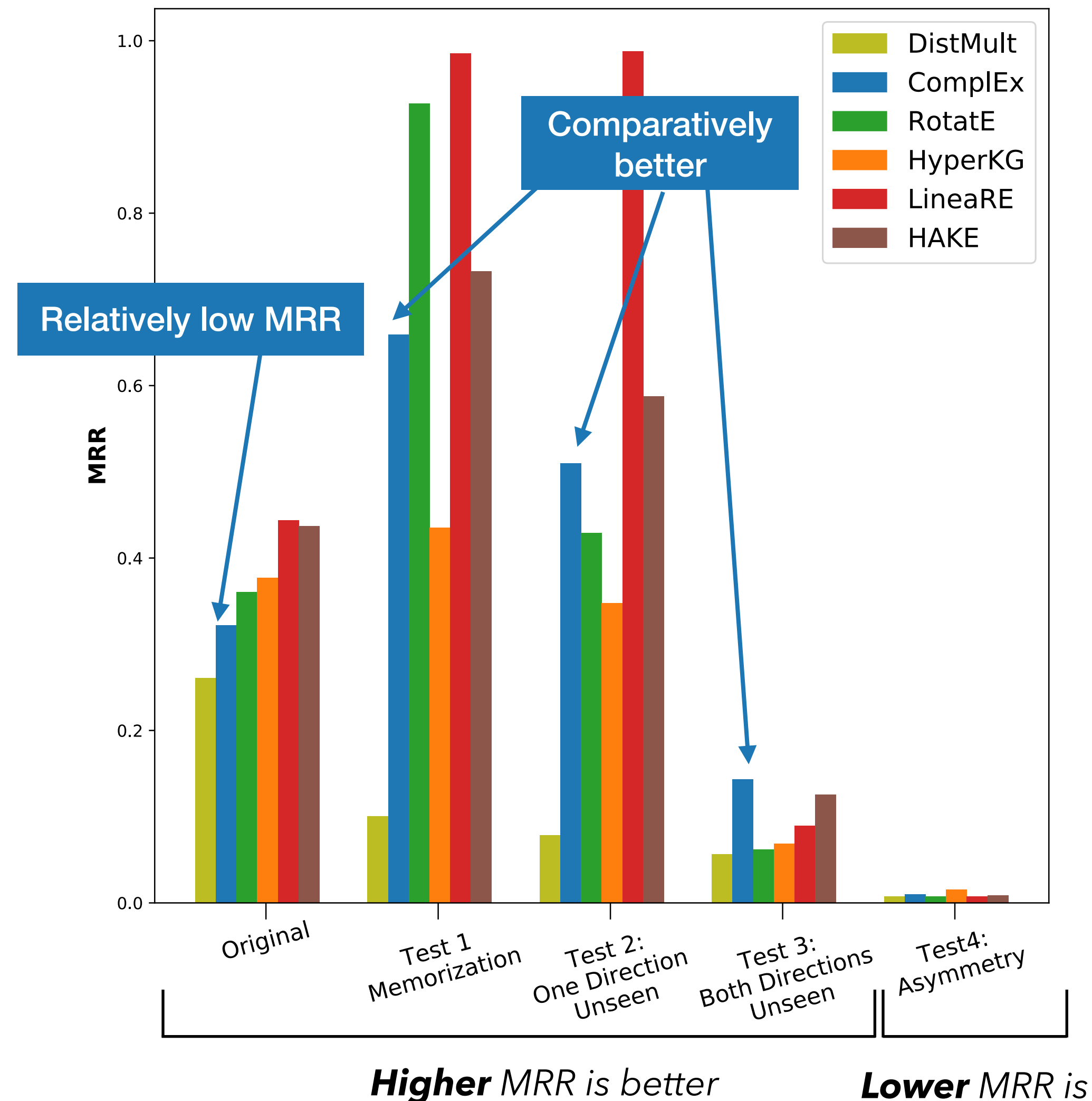
Proposed Method

Behavioral Testing for KGE models

For KGEs, we can target several different capabilities through different tests. We take the example of symmetric relations:



Results



A model can have a low average score but be the best at **generalizing on symmetry**.

KGEval: Understand the strengths and weaknesses of your link prediction models, source code at: <https://github.com/nec-research/KGEval>