We thank the reviewers for their very helpful comments and insights. We have grouped the actions we have taken to address these comments into three categories: methodological issues, issues with experiments or evaluations, and presentational issues. Some of the methodological issues raised by the reviewers arose due to miscommunications in the report – we have addressed these in the current version through changes in the text and figures, and we present these here under "Methodological>Clarifications".

Methodological

Changes

We thank the reviewers for noticing the inconsistency in negative distance loss. This loss function for distance was fixed by incorporating an indicator product to ensure zero loss when only one dimension was distinct. We retrained the models with the corrected loss, which marginally improved performance

Clarifications

Negative distance loss and disjointness

- Disjointness axioms were randomly selected, rather than using all available, for complexity reasons. The number of potential disjointness axioms scales quadratically with the number of classes, and this led to large memory and computational requirements. Finding a memory- and compute-efficient training regime for large graphs would be a valuable future contribution.

Influence of relational axioms on learning

- We clarified throughout the manuscript, both by changes to text and to figures, that our method does encode relational axioms, via the GNN. This means that all three of the axiom types mentioned influence the training of embeddings.

Other

- Symmetry of overlap loss function – this was a typographical error in the formula, which we have corrected.

Experimental/Evaluation

- We changed the example for demonstration from the family tree knowledge graph to the same knowledge graph used in the prediction task, focusing on presenting concepts from the "molecular function" domain. Some modifications were made to be able to train and show embeddings in two dimensions. We also included more detail on the evaluation of this method, for example that the hierarchy is not fully captured in the models. We also clarified the motivation for this learning task.

- Similarly for the link prediction evaluation, this was redone on the same knowledge graph. We also added more detail on the aim of this method and discussed related work. We also formalized the analysis of the distance distributions by performing Mann-Whitney U tests on the rank distributions of distances. The reviewers are right to point out that more evaluation is needed, which we raise in the results and discussion sections.
- We provide ablation studies justifying design choices made for the prediction models, including gene combination method, varying dimensionalities of embedding domains, and number of ignored (rare) edges.
- We thank the reviewers for their suggestions for alternative methods to compare against. However, we could not find the ProbE method mentioned despite searching the literature, and LogicE was not used as it is designed for embedding and answering complex queries rather than learning low dimensional KG representations. Instead we compared to baseline models based on ComplEx and Box^{2}EL embeddings.

Presentational

- We split the Introduction and Related works sections. The introduction now better introduces the problem and presents the structure of the paper. Related work is substantially expanded, giving more detail, especially, on KGE methods in general and box embedding methods in particular.
- We clarified and included additional detail on various concepts that were pointed out as insufficiently explained by the reviewers (e.g. GNN, GraphSAGE, Description Logics).
- We further separated general and application specific model descriptions, describing our general method in more detail in Section 5.1 and the task specific prediction models in Section 5.3.
- We give more detail on the hierarchies of the domains we split the KG into, such ontologies they were specified in, depth, and number of classes.
- The interaction hypothesis generation method was described more in detail, giving a better justification why it could correspond to a scientific hypothesis and a clearer description of how they were found
- Along with the ablation studies mentioned above, indicating improved performance, we give explain why we find it beneficial to vary and reduce the dimensionality of the domains.
- More detailed description is provided for Fig. 3 to make it more informative.
- Speaking terms have further been included in the Appendix where appropriate.
- The abstract has been updated to mention some results and with clarifications.
- We have updated the title of the manuscript to better reflect the content.
- Future research directions are discussed in Future work.