Appendix A. Joint Hierarchy Visualization

We qualitatively validate the performance of our box embedding method on modelling the ISA-HASPART joint hierarchy by visualizing the box embeddings for subsets of the joint hierarchy. One such subset occurring in WordNet is shown in figure 7. The box embeddings corresponding to this subset are shown in figures 8 and 9. It is worth noting that even with a classification F1 score of 68%, the visualizations validate our intuition.
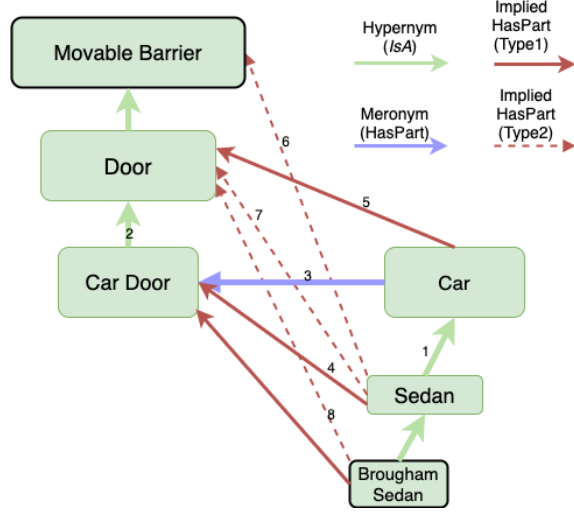

Figure 7: An example of joint hierarchy occurring in the WordNet (in the ISA and PARTOF relations)

The first parts of figures 8 and 9 show how the box embeddings are able to represent the implied HASPART edges – the inclusion of the Car and Sedan boxes inside the HASPART Car Door, HASPART Door and HASPART Movable Barrier boxes capture the edges 3,4 and 5 in figure 7. On the other hand, the second parts of the same figures show that, by inclusion of Sedan inside the Car box and Car Door inside the Door box, that our model still retains the edge information for edges 1 and 2.
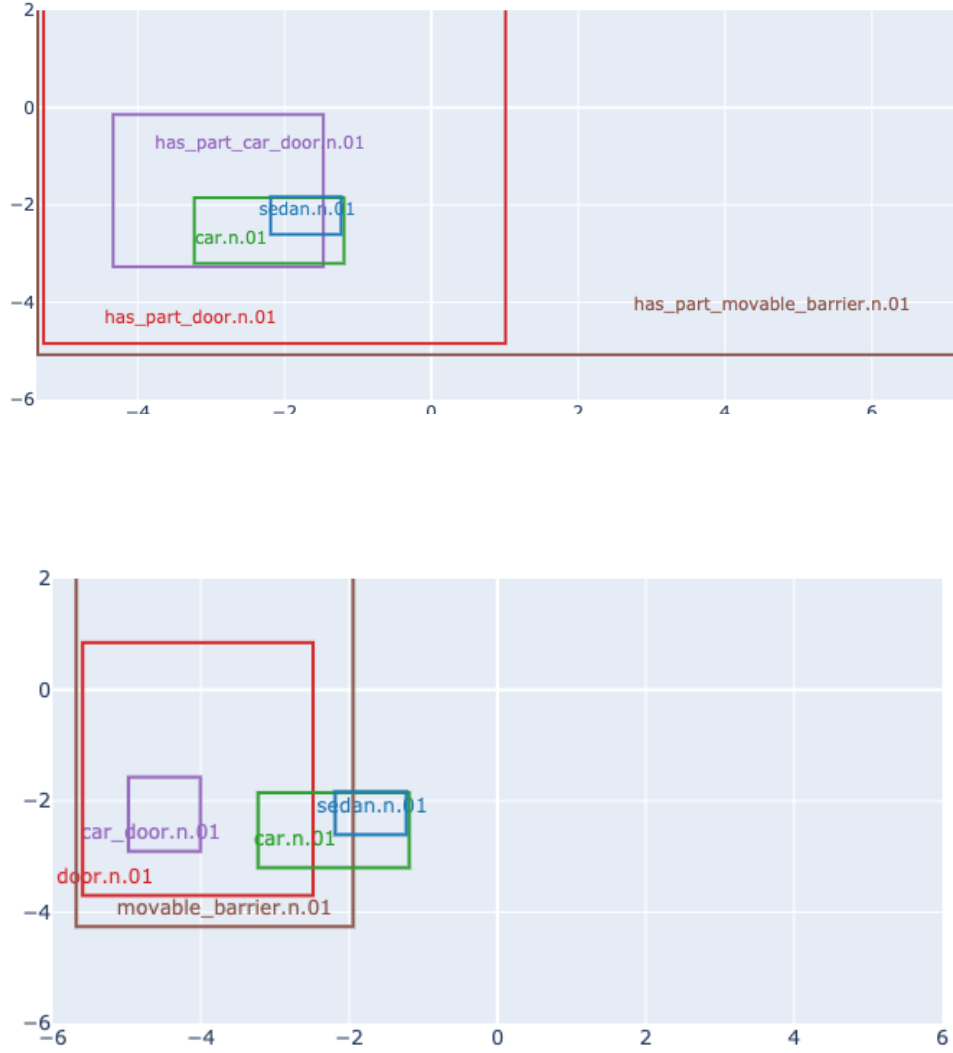
Figure 8: A 2-dimensional projection (on first two dimensions) of the 5-dimensional box embedding model trained to model the joint hierarchy. The embeddings plotted here correspond to the entities shown in figure 7. The two pictures show box embeddings in the same space but are separated into two plots for ease of visualization and comprehension. Note, some of the larger boxes extending outside of the plot range are not shown completely.
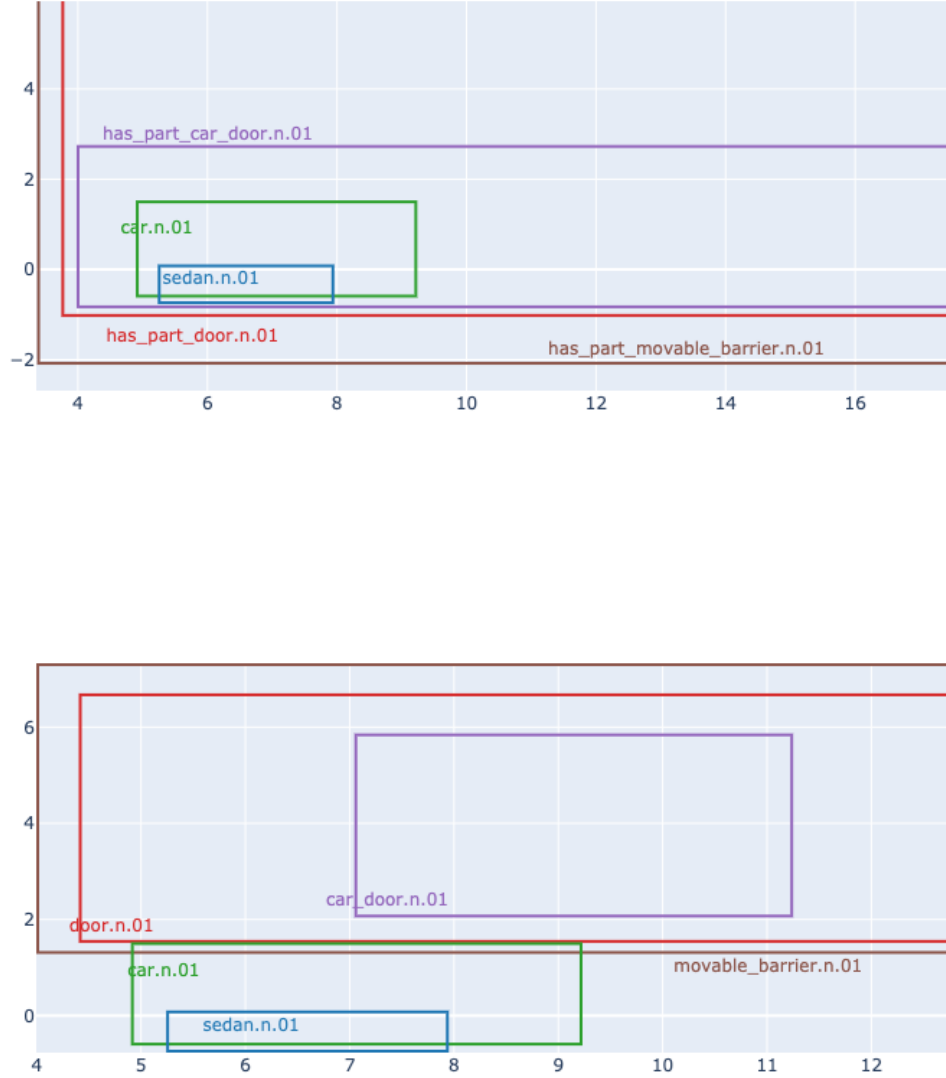
.

Figure 9: A 2-dimensional projection (on dimensions 3 and 4) of the 5-dimensional box embedding model trained to model the joint hierarchy. The embeddings plotted here correspond to the entities shown in figure 7. The two pictures show box embeddings in the same space but are separated into two plots for ease of visualization and comprehension. Note, some of the larger boxes extending outside of the plot range are not shown completely.

.

Appendix B. Effect of dimensions

We also investigate how different models scale with the number of parameters. As shown in table 4, the performance of all the models increase rapidly with the number of parameters, but plateaus after a certain point. This shows that all the models including box embeddings have strong inductive bias for modelling hierarchies. That is, they can model large hierarchical graphs with a small number of parameters.

Model	Test F1 score			
Vector Dimension (parameters per node)	10	20	50	100
Order Embedding	43	49.2	52.2	53
Poincare	28.9	30.9	31.2	31.4
Hyperbolic entailment cones	32.2	33.2	38.5	39.4
Box Embedding (our method)	60	65	67	68

Table 4: Effect of increasing model dimensions on the test F1 scores for the task of learning the Hypernym relation (with 0% transitive closure edges in the training set).