Appendices

A Negative Sampling

We train CoPER by minimizing the binary cross-entropy between for predicted distribution over answer entities. However, one challenge is that the training data only contains triples of the form (e_s, r, e_t) and there is no clear way to obtain negative samples of the form "this entity is not a correct answer to this question." A common approach to address this challenge is to consider all answers that appear in the training set as correct answers and all other entities in the KG as wrong answers. However, this approach exhibits two issues. First, it can become very expensive for large KGs as all the entities are involved in the computation of the loss function for each training example. In addition to that, it is not necessarily correct because some of the entities treated as incorrect answers during training are the very answers we are asked to infer at test time. To alleviate these two issues, we use an alternative approach based on that of Bordes et al. (2013) and Yang et al. (2015). Instead of considering all possible alternative answers as wrong, we uniformly sample a fixed number of alternative entities, for each positive triple, and use them as negative training examples. Our experiments show that this helps boost performance and also significantly improve efficiency.

B Similar Work Comparison Evaluations

In this section we show additional comparisons between CoPER, ConvE, MINERVA and two methods that allow for multiplicative interactions: TransR (Lin et al. 2015) and TransD (Ji et al. 2015). Further details on the architectures of these models are described in our Related Work Section. The results for TransR and TransD are as reported at https://github.com/thunlp/OpenKE, and all numbers represent `Hits@1`. We note that results for CTransR (Lin et al. 2015) are unavailable due to its unmaintained[1] or unavailable[2] implementation. While CoPER-MINERVA performs similarly to TransR, CoPER-ConvE significantly outperforms all other models on both datasets.

Model	Dataset	
	WN18RR	**FB15k237**
TransR	51.9	51.1
TransD	50.8	48.7
CoPER-ConvE	**56.12**	**62.97**
CoPER-MINERVA	50.99	50.39
ConvE	52.27	60.83
MINERVA	51.3	56.4

Table 1: Overview of `Hits@1` comparisons between CoPER models, TransR and TransD on FB15k237 and WN18RR.

[1] Its original implementation: https://github.com/thunlp/KB2E is no longer maintained and we were unable to train.

[2] It is missing from the official repository: https://github.com/thunlp/OpenKE.

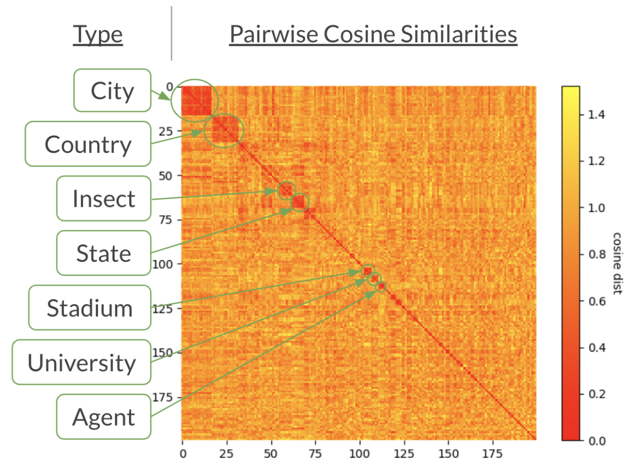

Figure 1: Heatmap of pairwise cosine similarities between relation embeddings in NELL-995. Relations are grouped according to manually defined "types" which we have gathered from analyzing the data. Each block corresponds to a section denoted by these groups.

C Relation Similarity Visualizations

Recall that one of our purported benefits of the contextual parameter generators described in our Parameter Generator Network Section was that relation information could be shared through the generator using approaches such as g_{linear}, enabling relations to leverage knowledge from their similar counterparts. To illustrate this, we analyze the resultant relation similarities after training CoPER-ConvE with g_{linear} on both FB15k-237 and NELL-995. Figure 2 showcases the relation similarity clusters formed by FB15k-237's relations. From the TSNE plot, we observe that many distinct relation clusters form through training. Moreover, among the 7 labeled clusters we randomly choose, we observe that their relations are all semantically similar to one another. Similarly, Figure 1 illustrates several observed relation similarity groupings. Specifically, the plot visualizes the pairwise similarities between each of the 200 relations in NELL, color-coded by their cosine distance from one another. Before plotting, each relation is grouped together according to manually defined "types" based on our analysis of the data. Based on the heatmap we observe several clear block diagonal structures, which indicate that relations are more similar to those of the same type than others from different types. For ease of understanding, we have labeled several blocks with their respective relation group "type". All materials relevant to our "type" assignment and code for this visualization can be found in our repository at `https://github.com/otiliastr/coper`. In addition, we display the top six most and least similar NELL-995 relations according to their cosine distance in Tables 2 and 3 respectively. Comparable to our findings in FB15k-237, we observe that the most similar relations convey comparable semantic information, while the least similar describe disparate data.

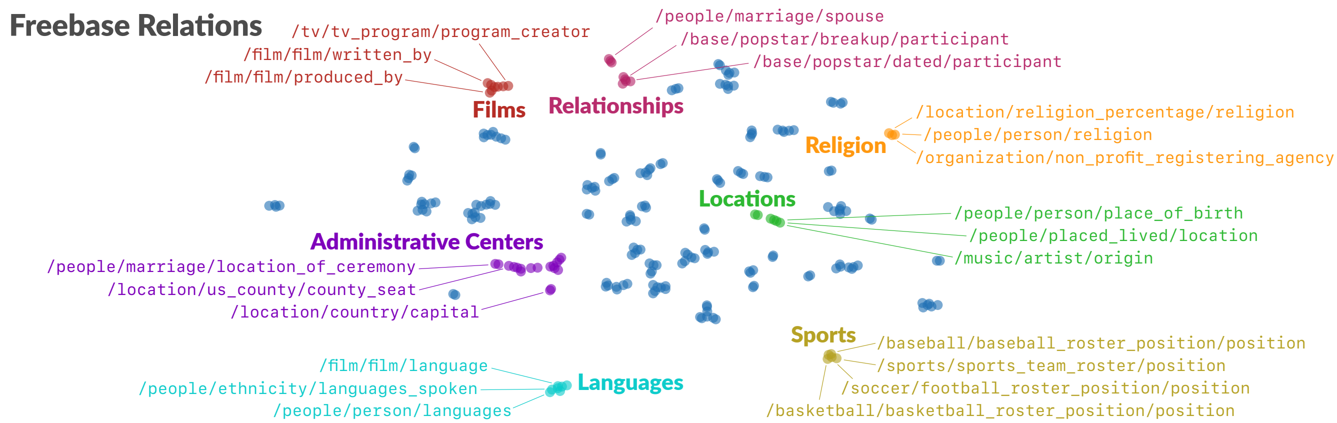

Figure 2: t-SNE visualization of FB15k-237 relations. We randomly label 7 clusters.

Relation Pair	Cosine Distance
(topmemberoforganization, ceoof)	0.01
(citylocatedincountry, citycapitalofcountry)	0.01
(organizationheadquarteredincity, radiostationincity)	0.01
(airportincity, buildinglocatedincity)	0.02
(athleteplaysforteam, athleteplaysforteam)	0.03
(organizationheadquarteredincity, televisionstationincity)	0.05

Table 2: Top six most similar pairwise relations measured by cosine distance in NELL-995.

Relation Pair	Cosine Distance
(synonymfor, sportsgamesport)	1.52
(sportschoolincountry, countrycurrency)	1.49
(synonymfor, teamplayssport)	1.48
(countrylocatedingeopoliticallocation, agentinvolvedwithitem)	1.47
(teamalsoknownas, athletewinsawardtrophytournament)	1.47
(statelocatedincountry, teamalsoknownas)	1.47

Table 3: Top six least similar pairwise relations measured by cosine distance in NELL-995.

References

Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, 2787–2795.

Ji, G.; He, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 687–696.

Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, 2181–2187. AAAI Press.

Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations (ICLR)*.