## A Datasets

FB15K (Bordes et al., 2013) is a dataset derived from Freebase. FB15K-237 (Toutanova et al., 2015) is a subset of FB15K which only contains the most frequent 237 relations, and where the inverse relations are removed to prevent test leakage. YAGO3-10 (Dettmers et al., 2018) is a dataset derived from YAGO-3 (Suchanek et al., 2007), where each entity occurs with at least 10 relations. FB1.9M is a large-scale dataset we have constructed from FB3M (Xu and Barbosa, 2018), a large dataset derived from Freebase by iteratively removing entities that occur in less than 5 triples until no such entities remain. The statistics for each of these datasets could be found in Table 1.

## B Implementation details

We optimize all models with stochastic gradient descent using Adam (Kingma and Ba, 2014), and perform early stopping with hits@10 on the validation set, where evaluation is performed every five epochs. For all our experiments, we fix initial learning rate to 0.001.

For experiments on FB15K, FB15K-237 and YAGO3-10, we fix embedding size to be 200. While performance increases with embeddings up to 2000 dimensions (Kadlec et al., 2017; Lacroix et al., 2018), we cap ours at 200 to emulate constraints faced when dealing with very large KGs. We perform a grid search over batch sizes: $\{500, 1000\}$, negative ratios $n_{neg}$ : $\{50, 100\}$, pair-loss weight $\alpha$ : $\{0.5, 1\}$ where applicable, and fix biased sampling probability $p$ to 0.3. We choose the hyperparameters that give the highest hits@10 on the validation set, and use these hyperparameters to report the final results on the test set.

For FB1.9M, we use the best hyperparameters from YAGO3-10. Due to memory constraints, we set the embedding size to 100.

For the ablation study on YAGO3-10, we perform a grid search over batch sizes: $\{200, 500, 1000\}$, negative ratios $n_{neg}$ : $\{50, 100\}$, biased sampling probability $p$ : $\{0.1, 0.2, 0.3\}$ and pair-loss weight $\alpha : \{0.25, 0.5\}$ where applicable.

For demonstrating how the effects of JoBi compares to baselines with varying batch sizes, we keep everything but the batch size constant ($n_{neg} = 25$, $\alpha = 0.5$, $p = 0.3$) and plot the change in hits@10 as the batch size varies in $\{25, 50, 100, 200, 500, 1000\}$. For demonstrating the effects of varying negative ratios, we keep everything but $n_{\text{neg}}$ constant (batch size = 200, $\alpha = 0.5$, $p = 0.3$) and plot hits@10 as $n_{\text{neg}}$ in $\{5, 10, 25, 50, 100, 200\}$.

## C Qualitative comparison between ComplEx and JoBi ComplEx

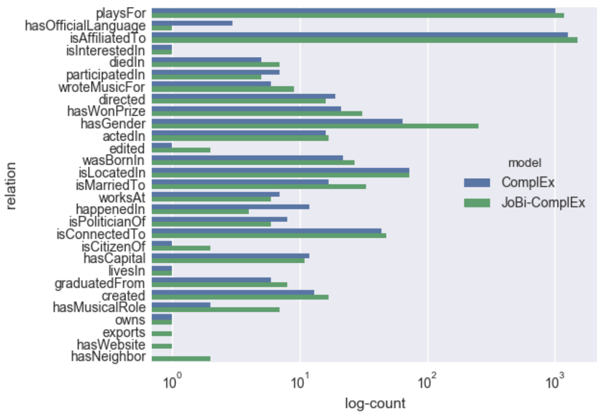


Figure 3: Number of correctly predicted entities (hits@1) for ComplEx and JoBi ComplEx, broken down by relation.

## D Experiments with full-softmax

Although the main focus of our framework is scalable training methods that use sampled negatives, we also test our joint method with softmax over the entire set of entities, and report the results in Table 7. We can see that joint training improves the performance also when used with full-softmax on all the datasets, beating a sophisticated, high performing method such as ConvE (Dettmers et al., 2018), and coming close to the performance of ComplEx-N3 (Lacroix et al., 2018) which uses embeddings 10 times larger in size than ours. We note that neither full-softmax, nor embedding sizes used by Lacroix et al. (2018) are scalable to large datasets.

## References

Ivana Balažević, Carl Allen, and Timothy M Hospedales. 2019. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th Interna-*

| FB15K-237 | Hits@1 | Hits@3 | Hits@10 | MRR |
|---|---|---|---|---|
| SimplE | 0.227 | 0.340 | 0.492 | 0.314 |
| DistMult | 0.225 | 0.343 | 0.490 | 0.313 |
| ComplEx | 0.229 | 0.348 | 0.502 | 0.319 |
| ConvE* | 0.237 | 0.356 | 0.501 | 0.325 |
| ComplEx-N3† | - | - | 0.56 | 0.37 |
| JoBi ComplEx | 0.238 | 0.357 | 0.509 | 0.327 |
| **FB15K** | Hits@1 | Hits@3 | Hits@10 | MRR |
| DistMult | 0.779 | 0.844 | 0.890 | 0.819 |
| ComplEx | 0.805 | 0.859 | 0.899 | 0.839 |
| ConvE* | 0.558 | 0.723 | 0.831 | 0.657 |
| ComplEx-N3† | - | - | 0.91 | 0.86 |
| JoBi ComplEx | 0.804 | 0.861 | 0.901 | 0.840 |
| **YAGO3-10** | Hits@1 | Hits@3 | Hits@10 | MRR |
| DistMult | 0.451 | 0.583 | 0.683 | 0.534 |
| ComplEx | 0.468 | 0.599 | 0.702 | 0.550 |
| ConvE* | 0.35 | 0.49 | 0.62 | 0.44 |
| ComplEx-N3† | - | - | 0.71 | 0.58 |
| JoBi ComplEx | 0.473 | 0.599 | 0.695 | 0.552 |

Table 7: Performance on different datasets against baselines and state-of-the-art methods using full-softmax. *Dettmers et al. (2018) †Lacroix et al. (2018)


*tional Conference on Neural Information Processing Systems*, pages 2787–2795. Curran Associates Inc.

Liwei Cai and William Yang Wang. 2018. Kbgan: Adversarial learning for knowledge graph embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1470–1480.

Kai-Wei Chang, Scott Wen-tau Yih, Bishan Yang, and Chris Meek. 2014. Typed tensor decomposition of knowledge bases for relation extraction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

Tim Dettmers, Minervini Pasquale, Stenetorp Pontus, and Sebastian Riedel. 2018. Convolutional 2D Knowledge Graph Embeddings. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence*, pages 1811–1818.

Alberto Garcia-Durán, Antoine Bordes, Nicolas Usunier, and Yves Grandvalet. 2016. Combining two and three-way embedding models for link prediction in knowledge bases. *Journal of Artificial Intelligence Research*, 55:715–742.

Prachi Jain, Pankaj Kumar, Soumen Chakrabarti, and others. 2018. Type-Sensitive Knowledge Base Inference Without Explicit Type Supervision. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 75–80.

Armand Joulin, Edouard Grave, Piotr Bojanowski, Maximilian Nickel, and Tomas Mikolov. 2017. Fast Linear Model for Knowledge Graph Embeddings. *arXiv preprint arXiv:1710.10881*.

Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. 2017. Knowledge Base Completion: Baselines Strike Back. *arXiv preprint arXiv:1705.10744*.

Seyed Mehran Kazemi and David Poole. 2018. SimplE Embedding for Link Prediction in Knowledge Graphs. *arXiv preprint arXiv:1802.04868*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Denis Krompaß, Stephan Baier, and Volker Tresp. 2015. Type-constrained representation learning in knowledge graphs. In *International Semantic Web Conference*, pages 640–655.

Timothe Lacroix, Nicolas Usunier, and Guillaume Obozinski. 2018. Canonical Tensor Decomposition for Knowledge Base Completion. In *Proceedings of the 35th International Conference on Machine Learning*.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2012. Factorizing YAGO: scalable machine learning for linked data. In *Proceedings of the 21st International Conference on World Wide Web*, pages 271–280, New York, New York, USA. ACM.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO. In *Proceedings of the 16th International Conference on World Wide Web - WWW '07*, page 697, New York, New York, USA. ACM Press.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing Text for Joint Embedding of Text and Knowledge Bases. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1499–1509, Lisbon. ACM.

Tho Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2071 – 2080.

Peng Xu and Denilson Barbosa. 2018. Investigations on Knowledge Base Embedding for Relation Prediction and Extraction. *arXiv preprint arXiv:1802.02114*.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *arXiv preprint arXiv:1412.6575*.