measure of performance, this should convincingly indicate that the top-k criterion is fundamentally inferior to the max-k criterion.

This is further comfirmed by comparing the F1 score of the sampling and greedy protocols against that of the **TopK_Oracle** limit. The F1 performances of all four models under these two protocols exceed **TopK_Oracle**, the theoretical upper limit that all possible models can ever achieve under the top-k criterion.

Besides, when applying the TopK protocol on the model's predictive distribution, we see that its F1 score is only slightly lower than the **TopK_Oracle** F1 curve and the gap between the two curves diminishes with k. That is, with the top-k criterion, there is little need to further improve the model, particularly at large k.

6.2.5 Comparing Models and Datasets Under The Max-K Criterion. Using a similar approach as that in Section 6.2.4, we now compare three state-of-the-art models, ComplEx, Analogy and ProjE, and assess the datasets with respect to these models.

The plots in Figure 4 are in the same form as those in Figure 3, except that we here only include the F1 metric. From the figure, one can conclude that under the F1 metric, ComplEx and Analog behave nearly identically. On FB15K, WN18 and WN18RR, they both outperform ProjE by a visible margin. On YAGO3-10, the three models have very similar performance. For each of the examined dataset in Figure 4, when comparing the model performance against the **MaxK_Oracle** limit, it is interesting to observe that on WN18, there is only very small gap between the **MaxK_Oracle** curve and the curves of these state-of-the-art models. That is, on this dataset, the performance of the current state of the art has already approached the ultimate limit. As such, we believe that on the WN18 dataset, further improvement upon these models will be very difficult, hence anticipating little progress in the coming years.

However, on FB15K, YAGO3-10 and particularly on WN18RR, the performance of the three models is still far away from the **MaxK_Oracle** limit. This suggests that there is still significant room for developing innovative models to further improve the performance on these datasets. One should not expect a model to achieve the **MaxK_Oracle** limit. To what extent a model can approach the **MaxK_Oracle** performance is in fact governed by the learnability bound (intrinsically dictated by the structure and size of the data), living somewhere below the **MaxK_Oracle** curve.

7 CONCLUDING REMARKS

Although the context of this work is link prediction in KBs, the max-k criterion and the protocols introduced in this paper in general apply widely to information retrieval, multi-label classification and many related areas. This criterion is practically motivated and at the same time theoretically sound. The proposed sampling and greedy protocols also have a theoretical foundation and a universal applicability. We anticipate that the max-k criterion and the proposed protocols find many applications beyond link prediction in KBs. Return to this context, we suggest that the conventional metrics used for link prediction be discarded, and replaced by the precision, recall and F1 metrics introduced herein.

ACKNOWLEDGMENTS

This work is supported partly by China 973 program (No.2014CB340305, 2015CB358700), by the National Natural Science Foundation of China (No. 61772059, 61602023, 61421003). This paper is also supported by the State Key Laboratory of Software Development Environment of China and Beijing Advanced Innovation Center for Big Data and Brain Computing.

REFERENCES

[1] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. *Dbpedia: A nucleus for a web of open data.* Springer.
[2] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data.* ACM, 1247–1250.
[3] Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. 2014. A semantic matching energy function for learning with multi-relational data. *Machine Learning* 94, 2 (2014), 233–259.
[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems.* 2787–2795.
[5] Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Conference on Artificial Intelligence.*
[6] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2017. Convolutional 2D Knowledge Graph Embeddings. *CoRR* abs/1707.01476 (2017). arXiv:1707.01476 http://arxiv.org/abs/1707.01476
[7] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2017. Convolutional 2D Knowledge Graph Embeddings. *CoRR* abs/1707.01476 (2017). arXiv:1707.01476 http://arxiv.org/abs/1707.01476
[8] G. E. Hinton, J. L. McClelland, and D. E. Rumelhart. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1. MIT Press, Cambridge, MA, USA, Chapter Distributed Representations, 77–109. http://dl.acm.org/citation.cfm?id=104279.104287
[9] Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. 2017. Knowledge Base Completion: Baselines Strike Back. In *Proceedings of the 2nd Workshop on Representation Learning for NLP, Rep4NLP@ACL 2017, Vancouver, Canada, August 3, 2017.* 69–74. https://aclanthology.info/papers/W17-2609/w17-2609
[10] Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2015. Modeling relation paths for representation learning of knowledge bases. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015* (2015).
[11] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of AAAI.*
[12] Hanxiao Liu, Yuexin Wu, and Yiming Yang. 2017. Analogical Inference for Multi-relational Embeddings. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. PMLR, International Convention Centre, Sydney, Australia, 2168–2178.
[13] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M. Suchanek. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias. In *CIDR 2015, Seventh Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 4-7, 2015, Online Proceedings.* http://cidrdb.org/cidr2015/Papers/CIDR15_Paper1.pdf
[14] Baoxu Shi and Tim Weninger. 2017. ProjE: Embedding Projection for Knowledge Graph Completion.
[15] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems.* 926–934.
[16] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web.* ACM, 697–706.
[17] Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality.* 57–66.
[18] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016.* 2071–2080. http://jmlr.org/proceedings/papers/v48/trouillon16.html
[19] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence.* Citeseer, 1112–1119.
[20] Jianfeng Wen, Jianxin Li, Yongyi Mao, Shini Chen, and Richong Zhang. 2016. On the Representation and Embedding of Knowledge Bases beyond Binary Relations. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016.* 1300–1307. http://www.ijcai.org/Abstract/16/188