FORSCHUNGSFORUM
DER ÖSTERREICHISCHEN
FACHHOCHSCHULEN
2026

CAMPUS 02 GRAZ
FACHHOCHSCHULE DER WIRTSCHAFT

Convexity for Trustworthy Knowledge Graph Embeddings: From Interpretability to Reliability

Hesham Morgan, TU Wien
Aleksandar Pavlović, University of Applied Sciences Campus Vienna
Ana Ozaki, University of Oslo, University of Bergen
Bruno F. Lourenço, The Institute of Statistical Mathematics
Emanuel Sallinger, TU Wien

Abstract: Knowledge graph embedding models infer missing links between entities, yet they are typically trained with gradient-based methods even when their latent structures are fundamentally convex. Convex optimization offers global optimality guarantees, efficient solvers, and transparent formulations defined by explicit constraints. We present a convex reformulation of BoxE that allows to express logical formulations as constraints in a convex optimization problem. The work we present is essential across multiple applications, such as drug development and autonomous vehicles, where ensuring safety-critical knowledge is paramount. Lastly, we evaluate our convex model on a subset of the Family dataset and demonstrate comparable link prediction performance.

Keywords: Knowledge graph embedding, Convex optimization, Interpretability, Reliability, Trustworthiness

1. Introduction

Knowledge graphs (KG) have become a central component of modern data-driven systems, supporting a wide range of tasks (Wang et al., 2017). They can be represented as a set of *entities* (also called *nodes*), *relations* (also called *edge labels*), and triples (also called edges or links) that connect two entities via a relation. An entity is the representation of a real-world object or concept, such as a person, country, or address (e.g. *JohnSmith* or *Austria*), while a relation corresponds to a type of relationship between two (or more) different entities (e.g. *livesIn*). Together, these components form statements, typically denoted as triples (*head entity*, *relation*, *tail entity*), such as (*JohnSmith*, *livesIn*, *Austria*). However, many downstream tasks rely on completing missing information of large KGs (Wang et al., 2017). This need for completion arises naturally: real-world KG data is inherently incomplete because of how it is collected, curated, and stored (Destandau & Fekete, 2021). KG embedding models (KGEs) address this need by (a) embedding entities and relations as objects in a latent space, (b) learning a scoring function based on these embeddings that shall score true triples high and false triples low, and (c) limiting the number of KGE parameters to be much smaller than those required to learn all true triples by heart, which forces the models to learn general patterns about the KG data and thereby allows not only to reconstruct the true KG triples but to infer missing ones.

Many KGEs, including BoxE (Abboud et al., 2020), ExpressivE (Pavlović & Sallinger, 2023a), and SpeedE (Pavlović & Sallinger, 2024b), ultimately use embeddings in the form of geometric regions. The use of such region-based embeddings offers a certain level of interpretability, allowing one to get a glimpse into which patterns (or logical rules) were learned by the model based on the spatial configuration of embeddings (Pavlović & Sallinger, 2023b; Pavlović & Sallinger, 2024a; Pavlović & Sallinger, 2025). However, while KGEs typically consider convex embedding regions, to date, the vast majority rely on stochastic gradient descent (SGD) optimization (Hogan et al., 2021). The lack of exploring convex optimization (CP) (Boyd & Vandenberghe, 2004) for the learning task is a missed opportunity, as CP offers several compelling advantages compared to gradient-based training. In convex optimization, the model acquisition procedure is explicitly defined by the objective function and constraints (Boyd & Vandenberghe, 2004), yielding a more transparent optimization problem whose feasible region is shaped directly by the mathematical formulations. This stands in contrast to gradient-based learning, where the training dynamics, implicit biases, and convergence behaviour can be more challenging to interpret (Chandrahas et al., 2020; Carvalho et al., 2019). Additionally, logical rules can be directly encoded in the convex model through the constraints on which the optimization problem is built. Moreover, convexity guarantees that every local optimum is a global optimum (Boyd & Vandenberghe, 2004), thereby removing ambiguity in the solution landscape and eliminating the risk of converging to poor local minima. Furthermore, there have been many studies on efficient convex optimization frameworks, such as CVXPY (Diamond & Boyd, 2016), which can also provide highly scalable performance, particularly on large but well-structured convex programs, often outperforming gradient-based methods in precision and stability (Mandi & Guns, 2020).

2. Application

These advantages of convex formulations make them particularly attractive in domains where interpretability, reliability, and reproducibility are essential. For example, in biomedical KGs, KGEs have been used to predict protein interactions and identify new drug candidates (Bonner et al., 2022). In such applications, understanding how a prediction is produced is crucial for clinical trust (Mustafa et al., 2025). Similar demands arise in political (Schwabe et al., 2019; Chen et al., 2017) and geopolitical settings (Gottschalk & Demidova, 2018), where high-stakes decisions from KGEs require transparent reasoning and predictable optimization behavior. Another application area is exploiting the information encoded in KGEs to improve the classification of decisions required for autonomous vehicle assistance and to predict the surrounding environment (Htun & Fukuda, 2024). Wickramarachchi et al. (2020) highlight the potential of KGEs to improve autonomous driving systems by facilitating neuro-symbolic systems. As noted by Htun et al. (2024), KGEs are also utilized to ensure compliance with safety and regulatory standards. In such settings, understanding how model representations are learned is vital, as downstream safety-critical decisions depend on their reliability. A further application area is cybersecurity, where embedding-based reasoning has been used to detect threats and identify countermeasures by analyzing attack patterns (Gilliard et al., 2024). The approach by Gilliard et al. (2024) aims to emulate human reasoning to derive actionable insights from security data. Our convex formulation could directly support this work, offering greater interpretability and the flexibility to encode domain-specific rules through tailored objectives and constraints, making it well-suited for precise, rule-driven cybersecurity tasks.

3. Contribution

Motivated by these considerations, we introduce a new convex reformulation of BoxE (Abboud et al., 2020), a classical KGE. We call our new formulation ConvBoxE, which enables efficient and more transparent embedding-based reasoning over KGs. ConvBoxE provides greater expressivity for inference involving constraints on entities and relations (e.g., cancer is a disease that genetic factors or lifestyle-related activities can cause), as well as concept disjointness constraints (e.g., smoking is not considered a direct causal factor for colon cancer).

ConvBoxE's formulation consists of several core components. Following the original BoxE framework by Abboud et al. (2020), entities (e.g., Cancer, Chemotherapy) are represented as vectors in a latent space, while relations (e.g., treatedBy) are represented as boxes. A triple (*Cancer*, *treatedBy*, *Chemotherapy*) is considered to be satisfied in ConvexBoxE if the embeddings of Cancer, Chemotherapy, and treatedBy satisfy a set of convex constraints. Intuitively, these constraints ensure that the points in the embedding space representing the combination of Cancer and Chemotherapy lie within the respective geometric regions (embedding boxes) associated with the relation treatedBy.

To learn such embeddings in ConvBoxE, we formulate the setting as a convex optimization problem. After the model is formalized, it is passed to a convex solver, which computes the optimal embeddings and relation boxes for the input dataset using an objective function. The objective function guides the solver toward an optimal solution and consists of multiple terms. In our case, the function includes terms for box membership discrepancy, center alignment, and negative sampling. The first two terms ensure that KG relations are properly learned and satisfied in the ConvBoxE embedding. To discourage unlikely triples from receiving high scores in the embedding space, negative sampling is incorporated by penalizing unlikely triples (e.g., cancer is caused by eating apples). While negative sampling is straightforward in SGD-based methods, it is challenging in convex optimization due to determinism and the non-convexity of the geometric complement of convex regions. ConvBoxE overcomes this by modeling negative sampling as a regularization term and by introducing convex complement regions for concept embeddings.

In addition to the objective function, our formulation includes a set of convex constraints that define the feasible region of the optimization problem. A key advantage of this formulation is the ability to incorporate domain knowledge reliably as explicit constraints in the optimization process. For example, in biomedical knowledge graphs, domain experts may know that certain drug combinations produce dangerous adverse interactions that must be avoided. Our framework can encode safety-critical knowledge as hard constraints, ensuring the learned embeddings respect them throughout training.

4. Evaluation

We have evaluated our ConvBoxE model on subsets of the Family dataset (Imenes et al., 2023), a well-established benchmark for testing the expressiveness of embedding models. It is ideal for assessing inference quality, as the dataset encodes structured patterns of family relations that enable meaningful evaluation even without specialized domain knowledge.

Next, we present preliminary results on a subset of the Family dataset. The subset contains 300, 110, and 440 triples for the training, validation, and testing datasets, respectively. Also, it contains 7 relations and 155 individuals. Although the dataset is rather small, it serves as a good testbed for our initial prototype and demonstrates the potential of our work. Additionally, the small size enables rapid experimentation and debugging during the prototyping phase. The experiments were conducted using an embedding dimensionality of 32, and the optimization problems were solved using MOSEK (MOSEK ApS, 2025) as the backend solver in CVXPY (Diamond & Boyd, 2016). The obtained results are presented in the table below.

Table 1:Evaluation of ConvBoxE and BoxE over the subset of the Family dataset

Metric	*ConvBoxE*	*BoxE*
MRR	0.8179	0.826
Hit@1	0.7022	0.719
Hit@3	0.9147	0.918
Hit@10	0.9806	0.986

The preliminary results for ConvBoxE and BoxE show that their performance is comparable. This finding shows that ConvBoxE already achieves strong predictive performance in our preliminary experiment, with further gains likely through refinements of the model's optimization formulation. Beyond accuracy, ConvBoxE also benefits from the theoretical and practical advantages of convex optimization, as discussed earlier, for reliably satisfying domain constraints in safety-critical tasks common in the pharmaceutical and autonomous driving sectors, among others.

Preliminary scalability tests also hint that for larger datasets, the solver might solve the resulting optimization problems within a reasonable time. As an example, we are currently evaluating ConvBoxE on a larger subset of the Family dataset (Imenes et al., 2023). The subset dataset contains 1061 nodes/entities and 5002 edges/triples. The generated convex program contains 69,249 variables and 142,479 constraints, yet CVXPY (Diamond & Boyd, 2016) required only 8.1 seconds to compile the problem. This hints at our approach being computationally feasible at larger scales and that convex formulations of KGE models likely remain practical as dataset sizes grow. We will investigate the scalability to larger real-world datasets in more detail in the future. The preliminary scalability experiments presented in this paper were conducted on a standalone computer with the following hardware specifications: a Ryzen 7 9700X CPU and 32 GB of DDR5 RAM operating at 6000 MHz. All computations reported so far were executed entirely on the CPU, without the use of GPU acceleration.

5. Conclusion

We demonstrate that convex formulations of KGEs offer improved transparency and the ability to guarantee domain constraints by directly formalizing them within the optimization problem, while preserving the interpretability provided by region-based models such as BoxE. These properties are particularly valuable in safety-critical domains such as biomedical and cybersecurity knowledge graphs, highlighting the promise of convex optimization as a foundation for more trustworthy and customizable embedding models.

In the future, we will focus on more efficient convex formulations of the problem and evaluate the proposed ConvBoxE model on larger, more complex datasets (such as FB15k-237 (Schlichtkrull et al., 2018) and WN18RR (Dettmers et al., 2018)). To better understand the potential advantages and limitations of convex approaches for KGs vs SGD-based approaches, we will systematically compare ConvBoxE's performance with that of more gradient-based models.

Acknowledgements

This work was funded by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/ICT25032, 10.47379/VRG18013, 10.47379/NXT22018, 10.47379/ICT2201, 10.47379/DCDH001], and by the Austrian Science Fund (FWF) 10.55776/COE12.

References

Abboud, R., Ceylan, U. U., Lukasiewicz, T., & Salvatori, T. (2020). BoxE: a box embedding model for knowledge base completion. Proceedings of the 34th International Conference on Neural Information Processing Systems. Presented at the Vancouver, BC, Canada. Red Hook, NY, USA: Curran Associates Inc.

Bonner, S., Barrett, I. P., Ye, C., Swiers, R., Engkvist, O., Hoyt, C. T., & Hamilton, W. L. (2022). Understanding the performance of knowledge graph embeddings in drug discovery. Artificial Intelligence in the Life Sciences, 2, 100036. doi:10.1016/j.ailsci.2022.100036

Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. Cambridge: Cambridge University Press.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics, 8(8), 832.https://doi.org/10.3390/electronics8080832

Chandrahas, Sengupta, T., Pragadeesh, C., & Talukdar, P. (2020, December). Inducing Interpretability in Knowledge Graph Embeddings. In P. Bhattacharyya, D. M. Sharma, & R. Sangal (Eds), Proceedings of the 17th International Conference on Natural Language Processing (ICON) (pp. 70–75). Retrieved from https://aclanthology.org/2020.icon-main.9/

Chen, W., Zhang, X., Wang, T., Yang, B., & Li, Y. (2017). Opinion-aware Knowledge Graph for Political Ideology Detection. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 3647–3653. doi:10.24963/ijcai.2017/510

Destandau, M., & Fekete, J.-D. (2021). The missing path: Analysing incompleteness in knowledge graphs. Information Visualization, 20(1), 66–82. doi:10.1177/1473871621991539

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2D knowledge graph embeddings. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11573

Diamond, S., & Boyd, S. (2016). CVXPY: a python-embedded modeling language for convex optimization. J. Mach. Learn. Res., 17(1), 2909–2913.

Gilliard, E., Liu, J., & Aliyu, A. A. (2024). Knowledge graph reasoning for cyber attack detection. IET Communications, 18(4), 297–308. doi:10.1049/cmu2.12736

Gottschalk, S., & Demidova, E. (2018). EventKG: A Multilingual Event-Centric Temporal Knowledge Graph. The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings, 272–287. Presented at the Heraklion, Greece. doi:10.1007/978-3-319-93417-4_18

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Maier, S. N., Ngomo, A. N., Polleres, A., Rashid, S., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge Graphs. ACM Computing Surveys, 54(4), 1–37. doi:10.1145/3447772

Htun, S. N. N., & Fukuda, K. (2024). Integrating Knowledge Graphs into Autonomous Vehicle Technologies: A Survey of Current State and Future Directions. Information, 15(10), 645. https://doi.org/10.3390/info15100645

Imenes, A., Guimarães, R., & Ozaki, A. (2023). Marrying Query Rewriting and Knowledge Graph Embeddings. Rules and Reasoning: 7th International Joint Conference, RuleML+RR 2023, Oslo, Norway, September 18–20, 2023, Proceedings, 126–140. Presented at the Oslo, Norway. doi:10.1007/978-3-031-45072-3_9

Mandi, J., & Guns, T. (2020). Interior Point Solving for LP-based prediction+optimisation. arXiv [Cs.NE]. Retrieved from http://arxiv.org/abs/2010.13943

MOSEK ApS. (2025). *MOSEK optimization toolbox for Python manual (Version* 11.0.3*)*. https://docs.mosek.com/

Mustafa, M., Al-Doghman, F., Hussain, F., Awadallah, M., & Ali, M. (2025). The role of explainable AI in knowledge graph-based drug repurposing: Bridging trust and discovery. In *2025 IEEE International Conference on E-Business Engineering (ICEBE)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICEBE68123.2025.00009

Pavlović, A. & Sallinger, E. (2023a). ExpressivE: A Spatio-Functional Embedding For Knowledge Graph Completion. In the 11th International Conference on Learning Representations, 11174. https://openreview.net/forum?id=xkev3_np08z

Pavlović, A. & Sallinger, E. (2023b). Building Bridges: Knowledge Graph Embeddings Respecting Logical Rules (short paper). In Proceedings of the 16th Alberto Mendelzon International Workshop on Foundations of Data Management. https://ceur-ws.org/Vol-3409/paper9.pdf

Pavlović, A. & Sallinger, E. (2024a). Raising the Efficiency of Knowledge Graph Embeddings While Respecting Logical Rules. In Proceedings of the 16th Alberto Mendelzon International Workshop on Foundations of Data Management. https://ceur-ws.org/Vol-3954/paper3573.pdf

Pavlović, A. & Sallinger, E. (2024b). SpeedE: Euclidean Geometric Knowledge Graph Embedding Strikes Back. In Findings of the Association for Computational Linguistics: NAACL 2024, 69–92. https://doi.org/10.18653/v1/2024.findings-naacl.6

Pavlović, A., Sallinger, E. & Schockaert S. (2025). Faithful Differentiable Reasoning with Reshuffled Region-based Embeddings. In Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning, 489-499. https://doi.org/10.24963/kr.2025/48

Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling Relational Data with Graph Convolutional Networks. The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings, 593–607. Presented at the Heraklion, Greece. doi:10.1007/978-3-319-93417-4_38

Schwabe, D., Laufer, C., & Busson, A. (2019). Building Knowledge Graphs About Political Agents in the Age of Misinformation. arXiv [Cs.CY]. Retrieved from http://arxiv.org/abs/1901.11408

Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge Graph Embedding: A Survey of Approaches and Applications. IEEE Transactions on Knowledge and Data Engineering, 29(12), 2724–2743. doi:10.1109/TKDE.2017.2754499.

Wickramarachchi, R.; Henson, C.A.; Sheth, A. An Evaluation of Knowledge Graph Embeddings for Autonomous Driving Data: Experience and Practice. In Proceedings of the AAAI Spring Symposium Combining Machine Learning with Knowledge Engineering, Palo Alto, CA, USA, 23–25 March 2020.