DORSET: Decoding hOw Knowledge Graph chaRacteristics Shape Embedding sTrategies (Invited Abstract)

Majlinda Llugiqi[1,*], Fajar J. Ekaputra[1,2], Marta Sabou[1], Nader Sobhi[2], Hesham Morgan[2], Eleonora Laurenza[2], Milos Jovanovik[2], Katja Hose[2] and Emanuel Sallinger[2]

[1]Vienna University of Economics and Business, Vienna, Austria
[2]TU Wien, Vienna, Austria

Abstract

AI applications across many domains increasingly draw on an emerging paradigm that combines two complementary approaches: symbolic and sub-symbolic AI. From the symbolic side, Knowledge Graphs (KGs) encode complex domain knowledge in formal graph structures. To make this knowledge usable by sub-symbolic methods that uncover latent patterns, KGs must first be embedded into low-dimensional vector spaces via KG Embeddings (KGEs), which serve as a bridge between symbolic structures and vector representations and are a pivotal component of modern AI systems. Nevertheless, a critical gap persists: how KG characteristics (e.g., expressivity) shape KGE performance remains poorly understood. So far, KGE research has relied on a small set of benchmarks with limited structural diversity, capping performance especially for specialized-domain KGs (e.g., medical KGs) with distinctive properties. The DORSET project (https://dorset-project.org/) aims to address this gap from two aspects: (1) building a systematic, experimentally grounded understanding of how KG characteristics shape KGE performance, and (2) using these insights to advance both KGE and knowledge engineering (KE) research. The project pursues this through iterative, cross-disciplinary collaborations between KGE and KE experts. Expected outcomes include diverse benchmarks, KG-characteristic-aware embedding methods (KGE), and new KG engineering methodologies and tools (KE), all with broader impact on AI applications of societal value.

Keywords

Knowledge Graph Embeddings, Knowledge Graph Characteristics, Knowledge Engineering

1. Introduction

Knowledge Graphs (KGs) have emerged as a key representation paradigm for knowledge-intensive domains such as biomedicine, finance, and engineering. By representing entities and their relationships in graph structures, KGs provide a semantically rich representation for integrating heterogeneous information and supporting reasoning and data-driven applications [1]. Knowledge Graph Embeddings (KGEs) complement this symbolic representation by transforming entities and relations into low-dimensional vector spaces, thereby making structured knowledge amenable to machine learning [2, 3]. KGEs support a broad range of downstream tasks, including link prediction, entity classification, clustering, recommendation, and knowledge discovery [4, 5]. Consequently, KGEs constitute an important bridge between symbolic and subsymbolic approaches to artificial intelligence and are increasingly used as components of neurosymbolic (NeSy) AI systems [6].

Despite this synergy, KGs and KGEs are largely studied as separate components. KGs exhibit substantial variation in their structural and semantic characteristics, including degree distributions, connectivity, relation types, schema constraints, semantic richness, temporal information, and topology.

Third Workshop on Knowledge Graphs and Neurosymbolic AI (KG-NeSy 2026), in ISWC 2026 Workshops Joint Proceedings, October 25–26, 2026, Bari, Italy

majlinda.llugiqi@wu.ac.at (M. Llugiqi); fajar.ekaputra@wu.ac.at (F. J. Ekaputra); marta.sabou@wu.ac.at (M. Sabou); nader.sobhi@tuwien.ac.at (N. Sobhi); hesham.morgan@tuwien.ac.at (H. Morgan); eleonora.laurenza@tuwien.ac.at (E. Laurenza); milos.jovanovik@tuwien.ac.at (M. Jovanovik); katja.hose@tuwien.ac.at (K. Hose); emanuel.sallinger@tuwien.ac.at (E. Sallinger)
0000-0002-5008-6856 (M. Llugiqi); 0000-0003-4569-2496 (F. J. Ekaputra); 0000-0001-9301-8418 (M. Sabou); 0009-0006-2403-4581 (H. Morgan); 0000-0002-2786-8163 (E. Laurenza); 0000-0001-7360-8015 (M. Jovanovik); 0000-0001-7025-8099 (K. Hose)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-05

CEUR-WS.org/Vol-4313/kgnesy-paper26.pdf

However, KGE research has predominantly relied on a small set of widely used benchmarks, such as FB15k, WN18RR, and YAGO3-10 [2, 7, 8, 9], which are often structurally unrealistic and biased [8, 10, 11]. As a result, it remains unclear how KG characteristics of more realistic KGs affect the performance of different KGE methods and downstream tasks.

This limitation is particularly relevant for specialised-domain KGs, whose characteristics may differ substantially from those of established benchmarks. Existing studies have reported effects of individual characteristics, such as node degree or inverse relations [12], on KGE performance, but these findings remain fragmented and are difficult to compare across datasets, models, and tasks. At the same time, knowledge engineering (KE) methodologies rarely consider the requirements of downstream KGE applications when constructing or adapting KGs. There is therefore a need for a systematic understanding of the impact of KG characteristics on KGE performance and for approaches that enable KGs and KGEs to be designed in synergy.

We argue that a deeper understanding of how KG characteristics shape KGE performance is essential for developing effective KG-based AI systems. We envision a shift from treating KGs as fixed inputs to embedding models towards a synergistic, profile-driven approach, in which KG characteristics, KGE methods, and downstream tasks are considered jointly. The DORSET project (https://dorset-project.org/) explores this vision through four complementary directions (work packages WPs): systematically investigating how KG characteristics shape KGE performance; developing KG profiles and diverse benchmarks; designing characteristic-aware KGE methods; and extending knowledge engineering methodologies to construct and adapt KGs for their intended applications.

2. Related Work

An increasing body of work demonstrates that characteristics of a KG can affect KGE performance. Several studies identify structural biases: high-degree or highly central entities tend to receive better predictions, while low-degree entities and infrequent entity–relation pairs are harder to learn [13, 8, 14, 15]. Other work shows that semantic KG characteristics also matter. For example, hierarchy depth affects the quality of embeddings [16], while semantic information such as lexical descriptions and logical constructors can improve ontology embedding and prediction tasks [17]. The effects of relation characteristics are similarly important but less consistent: inverse, duplicate, and Cartesian-product relations can improve link-prediction [10, 12] More recent studies further indicate that the impact of semantic information and KG characteristics can be model-dependent; for example, adding schema-level semantics benefits some embedding models while providing marginal or even negative effects for others [18, 19]. Beyond link prediction, such dependencies have also been observed for entity alignment, where density, heterogeneity, and negative sampling affect performance [20].

Despite these findings, the existing evidence remains fragmented across individual KG characteristics, KGs, KGE methods, and downstream tasks. Studies often investigate one characteristic in isolation and use different experimental settings, making it difficult to derive general principles about which KG characteristics matter, how they interact, and under which conditions they affect KGE performance. Moreover, commonly used benchmarks cover only a limited portion of the structural and semantic diversity of real-world KGs. This motivates a shift from evaluating KGE methods primarily across KGs towards understanding their performance across KG profiles—combinations of characteristics that capture relevant properties of a KG. DORSET builds on this emerging body of evidence to establish a systematic and experimentally grounded understanding of the KG–KGE interaction, and to use this understanding to inform both the development of characteristic-aware KGE methods and the engineering of KGs for downstream applications.

3. DORSET: Objectives and Research Agenda

DORSET is based on the hypothesis that a systematic understanding of the dependencies between KG characteristics and the performance of KGE-enabled tasks can provide the foundation for jointly

advancing KGE and KE. Its research agenda is organised around five objectives as shown in Figure 1.

- *O1: Understanding KG characteristics.* DORSET will systematically investigate how structural and semantic KG characteristics influence KGE performance across models and downstream tasks. This combines a systematic mapping of existing research with controlled experiments on real-world, synthetic, and systematically modified KGs.
- *O2/O3: KG profiling and benchmarks.* Based on these findings, DORSET will develop a framework for profiling KGs according to characteristics relevant to KGE performance and create benchmarks covering a broader range of KG profiles. Real-world KGs from domains including biomedicine, finance, and engineering will be complemented by synthetic and adapted KGs, enabling controlled investigation of individual characteristics and their combinations.
- *O4: Characteristic-aware KGE methods.* The resulting profiles and empirical evidence will guide the development of KGE methods that explicitly account for relevant KG characteristics, including topology, temporal information, schema and model heterogeneity, and latent knowledge. The goal is not to identify a universally best embedding model, but to develop robust and scalable approaches that are better suited to different KG profiles.
- *O5: Profile-based knowledge engineering.* Finally, DORSET will transfer the resulting insights back to KE by developing methodologies and tools for constructing and adapting KGs according to desired profiles. This includes controllable KG generation, profile-based KG adaptation, and the use of validation constraints such as SHACL. The resulting approaches will support a cyclic process in which KG engineering informs KGE method development, while KGE requirements feed back into KG engineering.

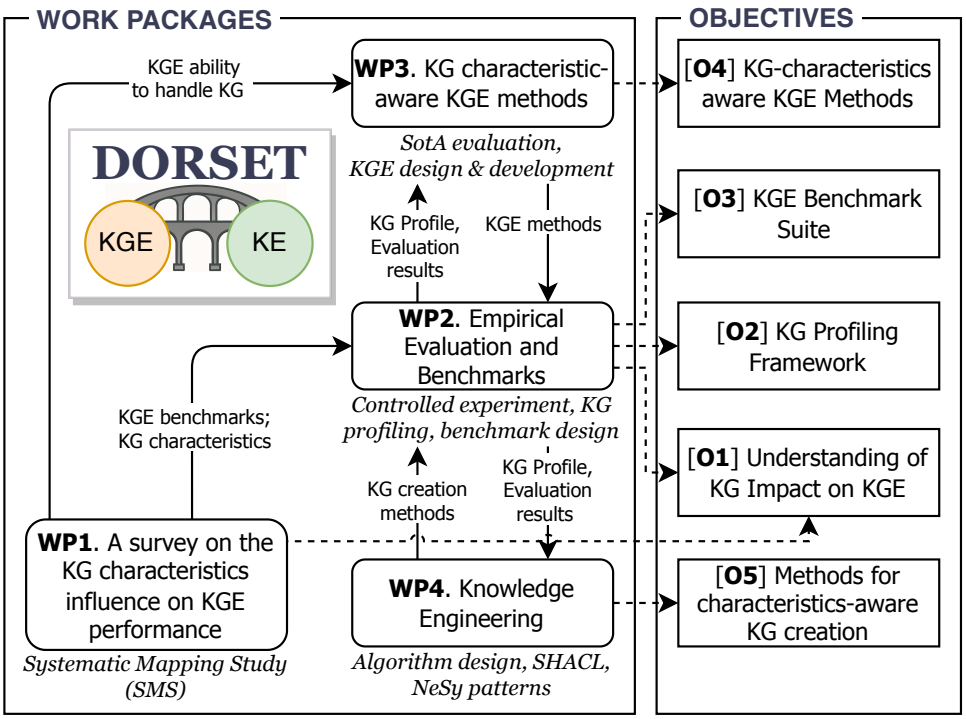


Figure 1: Overview of the work packages and objectives.

Together, these objectives establish a research loop connecting KG characteristics, KGE performance, KG profiles, KGE development, and KG engineering. This provides a basis for moving beyond benchmark-specific optimisation towards KG-aware and application-oriented development of embedding methods.

4. Conclusion

DORSET envisions a shift from treating KGs as static inputs to KGE methods towards jointly considering how knowledge is represented, how it is embedded, and what downstream tasks it is intended to support. By systematically studying KG characteristics, developing diverse benchmarks and profiling methods, creating characteristic-aware KGE models, and feeding these insights back into KE methodologies, DORSET aims to establish a synergistic KG–KGE research paradigm. Beyond advancing KGE and KE, this perspective has the potential to contribute to more robust NeSy AI systems and to improve KG-based applications in knowledge-intensive domains with high societal relevance.

Acknowledgments

This work is supported by DORSET Project (WWTF) 10.47379/ICT25032.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude and Grammarly in order to improve the language and readability of the text. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge Graphs, ACM Computing Surveys 54 (2022) 1–37. doi:10.1145/3447772.

[2] R. Biswas, L.-A. Kaffee, M. Cochez, S. Dumbrava, T. E. Jendal, M. Lissandrini, V. Lopez, E. L. Mencía, H. Paulheim, H. Sack, E. K. Vakaj, G. de Melo, Knowledge Graph Embeddings: Open Challenges and Opportunities, Transactions on Graph Data and Knowledge 1 (2023) 4:1–4:32. doi:10.4230/TGDK.1.1.4.

[3] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge Graph Embedding: A Survey of Approaches and Applications, IEEE Trans. on Knowl. and Data Eng. 29 (2017) 2724–2743. doi:10.1109/TKDE.2017.2754499.

[4] G. A. Gesese, R. Biswas, M. Alam, H. Sack, A survey on knowledge graph embeddings with literals: Which model links better literal-ly?, Semantic Web 12 (2021) 617–647. URL: https://doi.org/10.3233/SW-200404. doi:10.3233/SW-200404.

[5] D. Sellami, W. Inoubli, I. R. Farah, S. Aridhi, Knowledge graph representation learning: A comprehensive and experimental overview, Comput. Sci. Rev. 56 (2025) 100716. doi:10.1016/J.COSREV.2024.100716.

[6] P. Hitzler, A. Dalal, M. S. Mahdavinejad, S. S. Norouzi (Eds.), Handbook on Neurosymbolic AI and Knowledge Graphs, IOS Press, 2025. doi:10.3233/faia400.

[7] F. Bianchi, G. Rossiello, L. Costabello, M. Palmonari, P. Minervini, Knowledge graph embeddings and explainable AI, in: I. Tiddi, F. Lécué, P. Hitzler (Eds.), Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges, Studies on the Semantic Web, IOS Press, 2020, pp. 49–72. URL: https://doi.org/10.3233/SSW200011. doi:10.3233/SSW200011.

[8] A. Rossi, A. Matinata, Knowledge Graph Embeddings: Are Relation-Learning Models Learning Relations?, in: Proc. of the Workshops of the EDBT/ICDTJoint Conference, 2020.

[9] W. Zhang, J. Chen, J. Li, Z. Xu, J. Z. Pan, H. Chen, Knowledge graph reasoning with logics and embeddings: Survey and perspective, in: 2024 IEEE International Conference on Knowledge Graph, IEEE, 2024, pp. 492–499.

[10] F. Akrami, M. S. Saeef, Q. Zhang, W. Hu, C. Li, Realistic Re-evaluation of KG Completion Methods: An Experimental Study, in: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, 2020, pp. 1995–2010. doi:10.1145/3318464.3380599.

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

[12] M. Ali, M. Berrendorf, C. T. Hoyt, L. Vermue, M. Galkin, S. Sharifzadeh, A. Fischer, V. Tresp, J. Lehmann, Bringing Light Into the Dark: A Large-Scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2022) 8825–8845. doi:10.1109/TPAMI.2021.3124805.

[13] S. Bonner, U. Kirik, O. Engkvist, J. Tang, I. P. Barrett, Implications of topological imbalance for representation learning on biomedical knowledge graphs, Briefings Bioinform. 23 (2022). doi:10.1093/BIB/BBAC279.

[14] A. Sadeghi, D. Collarana, D. Graux, J. Lehmann, Embedding knowledge graphs attentive to positional and centrality qualities, in: Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference, ECML PKDD, volume 12976, Springer, 2021, pp. 548–564. doi:10.1007/978-3-030-86520-7_34.

[15] H. Shomer, W. Jin, W. Wang, J. Tang, Toward degree bias in embedding-based knowledge graph completion, in: Proceedings of the ACM Web Conference 2023, 2023, pp. 705–715.

[16] M. Ayoughi, M. van Spengler, P. Mettes, P. Groth, Designing hierarchies for optimal hyperbolic embedding, in: European Semantic Web Conference, Springer, 2025, pp. 362–382.

[17] J. Chen, P. Hu, E. Jimenez-Ruiz, O. M. Holter, D. Antonyrajah, I. Horrocks, Owl2vec*: Embedding of owl ontologies, volume 110, Springer, 2021, pp. 1813–1845.

[18] B. Dave, A. Christou, C. Shimizu, Towards understanding the impact of graph structure on knowledge graph embeddings, in: International Conference on Neural-Symbolic Learning and Reasoning, Springer, 2024, pp. 41–50.

[19] E. Harlander, Influence of knowledge graph characteristics on embedding-based recommender systems (2024).

[20] N. Fanourakis, V. Efthymiou, D. Kotzinos, V. Christophides, Knowledge graph embedding methods for entity alignment: experimental review, Data Mining and Knowledge Discovery 37 (2023) 2070–2137.