Semantic Similarity through Human-in-the-loop in Knowledge Graph Embeddings

Verdiana Schena[a,*]

Supervised by Prof. Simona Colucci and Prof. Floriano Scioscia
[a]Politecnico di Bari, Bari, Italy
ORCID (Verdiana Schena): https://orcid.org/0009-0009-1660-5462

Abstract. Knowledge Graph Embeddings are commonly used to represent logical resources as entities and relations in vector spaces. Their performance has been mainly demonstrated through the link prediction task. However, their effectiveness for other applications, such as semantic similarity, remains an open question. This doctoral project proposes the development, fine-tuning, and application of LISE (Logic-based Interactive Similarity Explainer) for evaluating Knowledge Graph Embeddings in tasks related to similarity. LISE is a modular system integrating heterogeneous approaches: (i) machine learning for Knowledge Graph Embeddings and clustering; (ii) logic-based reasoning to identify common characteristics among clustered resources; (iii) natural language generation to provide human-understandable explanations of similarities. A key feature of LISE is its human-in-the-loop module, which lets users provide feedback on the relevance of the explanations. This feedback is then used to refine the embedding and clustering processes, with the goal of moving the similarity results closer to the user expectations.

1 Introduction and Motivation

Knowledge Graph Embeddings (KGE) map entities and relations from a Knowledge Graph (KG) into low-dimensional vector spaces [3], aiming to capture their semantic relationships. Knowledge Graph Embedding Models (KGEMs) are primarily evaluated on the link prediction task, *i.e.*, predicting missing heads or tails in triples (head, relation, tail) [14], by employing metrics such as Mean Reciprocal Rank and HITS@K. KGEMs are also applied to other tasks including recommendation, knowledge base completion [3], entity similarity, and conceptual clustering [11]. However, in recent years, it is emerging that KGEMs are not always suitable for all of these applications [11, 12].

This research project aims at investigating the effectiveness of KGEs in the semantic similarity tasks, *i.e.*, to evaluate the degree of similarity between two Resource Description Framework (RDF) [9] resources.

To this purpose, we propose the use of LISE, a system that combines machine learning, knowledge-based reasoning, natural language generation, and interactive feedback. LISE offers a post-hoc and method-agnostic approach to explain the similarity of RDF resources and proposes the refinement of KGEs for interactive clustering based on user needs [7, 8]. This work addresses some remaining open issues through two *Research Questions* (RQs):

- **RQ1**: *How can semantic similarity in Knowledge Graph Embeddings be effectively evaluated beyond link prediction approaches?*
- **RQ2**: *Can user interactivity contribute to refine the distribution of resources in the embedding space, thereby aligning a suitable distance measure in the space more closely to user expectations?*

Section 2 provides an overview of the literature relevant to both RQs. Section 3 describes the approach under investigation in the Ph.D program. Section 4 presents the initial results and discusses future work for RQs.

2 Related Work

2.1 Semantic Similarity in Knowledge Graph Embeddings

The assumption that KGEMs capture the semantic similarity of entities and relations, mapping similar entities (or relations) to similar vectors [3], has recently been challenged by some studies [11, 12]. These works highlight the need to explore alternative metrics to those evaluating the performance of embeddings on the link prediction task, which currently dominates the field. Specifically, it is questioned whether semantically similar entities are actually represented by similar embeddings [11]. Moreover, there is no universally accepted definition of entity similarity, which remains ambiguous and context-dependent, as each KGEM may encode a distinct notion of similarity [11]. Embeddings similarity is usually assessed using metrics such as cosine similarity [3]. However, Hubert *et al.* [11] argue that the ranking-based metrics used for link prediction are poor indicators of semantic coherence in the embedding space. They propose alternative evaluation strategies considering, for example, that two entities might be similar if they share a set of predicates in the KG. In particular, entities sharing predicates with a set of other entities might be close in the embedding space [11].

2.2 Human-in-the-loop in Machine Learning

Miller [13] provides an understanding of the terms 'interpretability' (or 'explainability') and 'explanation'. The former is defined as the degree to which a human can understand the decision-making process within a given context; the latter is the explication of the reasoning behind those decisions to people. Explanations defined in this way have a twofold utility. First, they allow users to understand the internal functioning of machine learning models, often defined

* Corresponding Author. Email: v.schena@phd.poliba.it

as black boxes given their lack of transparency regarding decision-making [2, 13]. Second, these explanations address challenges in human-agent interaction [13]. In fact, going beyond the purely objective explanation of the internal operations of machine learning algorithms, their utility can include the necessity of a subjective interpretation, focusing on what the user specifically needs to understand about the AI. To achieve this, the human-in-the-loop approach enables human intervention within the AI's decision-making process [2]. Defined as a fundamental component of Interactive Machine Learning and *Symbiotic Artificial Intelligence* (SAI) [10], human-machine cooperation aims to achieve a specific task through collaboration between a human and a machine. While a combined approach has been shown to outperform fully automated or manual methods, the specific advantages of its application, as well as its objective and subjective metrics, require further study [4].

3 Proposed Approach: LISE

This doctoral project is the development, fine-tuning, and application of LISE [7, 8] (Logic-based Interactive Similarity Explainer): a system designed to explain similarities within clusters of RDF resources. Its architecture consists of four distinct modules.

LISE's ***Machine Learning Module*** generates KGE of RDF resources from a given KG and subsequently groups them using a clustering algorithm. The clusters' items are represented by numerical embeddings, making it difficult to understand the common properties shared by clustered resources and the reasons for their location in the embedding space. In this way, it remains unclear whether these embeddings truly capture the semantic meaning of entities, allowing similar resources to be grouped together in the embedding space. The only metrics derivable from this module are purely numerical similarity measures, such as the silhouette score and cosine similarity. While these measures offer insights into the cohesion of the clusters and provide a general overview of the resources' distribution, they do not clarify the common characteristics among clustered entities. To overcome this limitation, LISE integrates a ***Logic-based Module*** that processes the RDF KG to compute the Least Common Subsumer (LCS) [6]. For each cluster, the LCS generates an RDF KG that explicitly describes the common features of the RDF resources it contains. In this module, the LCS is refined by pruning information identified as irrelevant to the KG context or the user's knowledge [5], thereby becoming a Common Subsumer (CS). With the aim to make the CS human-readable, the system presents, in its ***Natural Language Generation Module***, a template-based approach that generates clear and structured explanations of each cluster's commonalities. Finally, in LISE's ***User Interaction and Feedback Loop Module***, these explanations are rated by users according to the relevance of information by using a star rating system. This module aims to collect the user's ratings, train a regression model, and predict weights used afterwards to refine the embedding process and the clustering approach, thereby leveraging user needs. This module is fundamental for the interactivity of clustering; the feedback loop is an essential phase since the notion of similarity is subjective and depends on the individual user's objectives [1]. By integrating user feedback, LISE adapts the embeddings, the clustering, and, thus, the generated explanations to better align with the user's perspective.

4 Results and Future Work

So far [7, 8] LISE has been evaluated using the *Drugbank*[1] dataset. This evaluation has employed *pyRDF2Vec*[2] as KGEM and *k-means* as clustering algorithm. Despite tuning some hyperparameters for both embedding and clustering models, the results in terms of the shared characteristics among clusters are not impactful, as limited and few informative commonalities are identified. This has led the investigation to align with the recent literature [11, 12], questioning whether KGEMs are effective only for the link prediction task, or if they can also be significant for other tasks like clustering or entity similarity. To better explore the question defined in **RQ1**, future work aims to:

- Compare the performance of KGEMs on the link prediction task and semantic similarity tasks;
- Compare different KGEMs for similarity tasks, identifying models better capturing the semantics of resources;
- Identify useful metrics that go beyond numerical distances among resources in the embedding space and possibly consider their semantics.

To refine the distribution of resources within the embedding space, thereby enhancing the alignment of similarity results with the user expectation (**RQ2**), LISE currently uses a custom sampling strategy within pyRDF2vec, *Predicate Relevance Weight* [7, 8]. This strategy assigns higher weights to predicates that appear in sentences users consider more relevant. However, the strategy's ability to impact embedding significance has not yet been fully explored, due to the limited number of user-provided relevance votes. In our preliminary experiments, we tested this approach by assigning very low weights to certain predicates and higher weights to others, allowing the model to prioritize some predicates over others during the embedding process. To enable users to assign these ratings, preliminary explorations regarding the verbalization of CS with either template-based explanations, or direct use of an LLM (*e.g.,* Google Gemini), or a combination of both, have already been conducted.

For **RQ2**, ongoing work is focused on two main directions:

- Collecting a sufficient number of user-assigned weights to train a regression model and subsequently retrain the embedding model using the custom sampling strategy based on user relevance feedback;
- Investigating how this weighting mechanism can be extended and applied to KGEMs different from pyRDF2Vec.

Acknowledgements

This work is partially funded by the Italian Aerospace Technological District 'Distretto Tecnologico Aerospaziale S.C.a.r.l.'. The candidate thanks Francesco M. Donini for useful discussions and collaboration on the research project. The author also gratefully acknowledges the support of EurAI through the ECAI Travel Grant.

References

[1] J. Bae, T. Helldin, M. Riveiro, S. Nowaczyk, M.-R. Bouguelia, and G. Falkman. Interactive clustering: A comprehensive review. *ACM Computing Surveys (CSUR)*, 53(1):1–39, 2020.

[1] https://download.bio2rdf.org/files/current/drugbank/drugbank.html
[2] https://pyrdf2vec.readthedocs.io/en/latest/index.html

[2] I. Baroni, G. Re Calegari, D. Scandolari, and I. Celino. AI-TAM: a model to investigate user acceptance and collaborative intention in human-in-the-loop AI applications. *Human Computation*, 9(1):1–21, May 2022. doi: 10.15346/hc.v9i1.134.

[3] F. Bianchi, G. Rossiello, L. Costabello, M. Palmonari, and P. Minervini. *Knowledge Graph Embeddings and Explainable AI*. IOS Press, 2020. doi: 10.3233/ssw200011. URL http://dx.doi.org/10.3233/SSW200011.

[4] I. Celino. Who is this explanation for? Human intelligence and knowledge graphs for eXplainable AI. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, pages 276–285. IOS Press, 2020.

[5] S. Colucci, F. M. Donini, and E. Di Sciascio. On the relevance of explanation for RDF resources similarity. In *Proc. of 3rd International workshop on Model-driven Organizational and Business Agility, MOBA 2023*, volume 488 of *Lecture Notes in Business Information Processing*, pages 96–107. Springer, 2023.

[6] S. Colucci, F. M. Donini, and E. Di Sciascio. Computing the Commonalities of Clusters in Resource Description Framework: Computational Aspects. *Data*, 9(10), 2024. ISSN 2306-5729. doi: 10.3390/data9100121.

[7] S. Colucci, F. M. Donini, and V. Schena. LISE: a Logic-based Interactive Similarity Explainer. In *SeWebMeDA-2025: 8th International Workshop on Semantic Web Solutions for Large-scale Biomedical Data Analytics*, 2025. In press.

[8] S. Colucci, F. M. Donini, V. Schena, F. Scioscia, and E. Di Sciascio. LISE: A Logic-Based Interactive Similarity Explainer for Clusters of RDF Data. *IEEE Access*, 13:90109–90128, 2025. doi: 10.1109/ACCESS.2025.3571518.

[9] R. Cyganiak, D. Wood, and M. Lanthaler. Rdf 1.1 concepts and abstract syntax, w3c recommendation, 2014. URL https://www.w3.org/TR/rdf11-concepts/.

[10] G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, and M. F. Costabile. From human-centered to symbiotic artificial intelligence: a focus on medical applications. *Multimedia Tools and Applications*, pages 1–42, 2024.

[11] N. Hubert, H. Paulheim, A. Brun, and D. Monticolo. Do Similar Entities Have Similar Embeddings? In A. Meroño Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, and P. Lisena, editors, *The Semantic Web*, pages 3–21, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-60626-7.

[12] N. Jain, J.-C. Kalo, W.-T. Balke, and R. Krestel. Do Embeddings Actually Capture Knowledge Graph Semantics? In R. Verborgh, K. Hose, H. Paulheim, P.-A. Champin, M. Maleshkova, O. Corcho, P. Ristoski, and M. Alam, editors, *The Semantic Web*, pages 143–159, Cham, 2021. Springer International Publishing. ISBN 978-3-030-77385-4.

[13] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2018.07.007.

[14] M. Pote. Survey on Embedding Models for Knowledge Graph and its Applications, 2024. URL https://arxiv.org/abs/2404.09167.