Bringing Back Semantics to Knowledge Graph Embeddings : An Interpretability Approach

Antoine Domingues[1,2], Nitisha Jain[1], Albert Meroño Peñuela[1], and Elena Simperl[1]

[1] Department of Informatics, King's College London, London, UK
[2] ENSTA Paris, Institut Polytechnique de Paris, France
{antoine.domingues, nitisha.jain}@kcl.ac.uk

Abstract. Knowledge Graph Embeddings Models project entities and relations from Knowledge Graphs into a vector space. Despite their widespread application, concerns persist about the ability of these models to capture entity similarity effectively. To address this, we introduce *InterpretE*, a novel neuro-symbolic approach to derive interpretable vector spaces with human-understandable dimensions in terms of the features of the entities. We demonstrate the efficacy of *InterpretE* in encapsulating desired semantic features, presenting evaluations both in the vector space as well as in terms of semantic similarity measurements.

Keywords: knowledge graph embeddings · semantic similarity · interpretable vectors.

1 Introduction

Since early 2010s, significant progress has been made in the development of Knowledge Graph Embeddings Models (KGEMs). These models aim to project the entities and relations of Knowledge Graphs (KGs) in a latent vector space. This approach offers a sub-symbolic means of representing the entities and their connections within the original KG [3]. KGE models have found applications across various tasks, including KG completion, rule-based reasoning, and recommendation systems [25, 11]. These models are typically trained and evaluated with a focus on the task of link prediction where a score for plausibility of KG triples is optimized.

However, there is a prevalent belief that KGEMs can effectively capture similarities between underlying entities where similar entities are clustered in the vector space. As such, KGEMs have been used for tasks such as entity or relation similarity and conceptual clustering [16, 9, 21]. This notion was first challenged by Jain et al. [14], where the authors demonstrated that entities belonging to the same type (or ontological class) do not effectively cluster together in the vector space beyond the most basic entity types. Subsequently, other recent studies have delved into this further, arriving at similar conclusions [12, 1].

A fundamental challenge for KGEMs in terms of capturing entity similarity stems from the complex nature of the underlying data. Entities within the

KG possess diverse features, such as attributes and relations to other entities, which significantly influence their vector representations. This complexity makes it exceedingly difficult to pinpoint the precise factors that shape the distribution of vectors in the embedding space. With the entities having varying types and numbers of connections in the KG, and learned vector representations consisting of hundreds of dimensions, there exists no direct correspondence between the entity features and the dimensions of the resulting vector. The lack of mapping leads to a deficiency in semantic interpretability, with no means to comprehend why certain vectors are similar, nor to identify which entity features influence the representations.

In this work, we aim to bring back the semantic interpretability for the embedding vectors by explicitly connecting them to underlying features of the entities. Our proposed neuro-symbolic approach *InterpretE* is capable of deriving new vector spaces that can be understood in terms of the human-understandable features of the entities in the KG, hence enabling informed decisions in downstream semantic tasks (e.g. recommendation systems and conceptual clustering), debugging and comparing the models and understanding hidden biases [20]. We design different experiments to demonstrate that the vector spaces obtained from *InterpretE* can encapsulate desired semantic features and the approach is highly flexible in terms of the number and types of the entity features. The evaluation of the approach is presented in terms of the quality of the resulting clusters in the derived vector space, as well in terms of the semantic similarity of the corresponding entities. We make the code publicly available [3] to promote further research in this direction.

2 Related work

Semantics in Knowledge Graph Embeddings. Recent critiques have questioned the widely-held assumption that KGEMs produce semantically meaningful representations of underlying entities [14, 12]. In a popular previous work, Jain et al. [14] investigated the degree to which similar entities correspond to similar vectors and concluded that this does not hold true universally. They demonstrated that entity embeddings derived from KGEMs often struggle to effectively discern entity types within a Knowledge Graph (KG), with simpler statistical methods offering comparable performance. Additionally, Ilievski et al. [13] observed consistent under-performance of KGEMs compared to simpler heuristics in tasks reliant on similarity, particularly within word embeddings. The authors argue that many properties that heavily relied upon by KGEMs are not conducive to determining similarity, thereby introducing noise that ultimately undermines performance.

Interpretable Dimensions. Several approaches have emerged to construct interpretable spaces [7, 5, 4, 20, 27] using multiple data sources, predominantly texts but also images. The term 'interpretable space' encompasses simple and

[3] `https://github.com/toniodo/InterpretE`

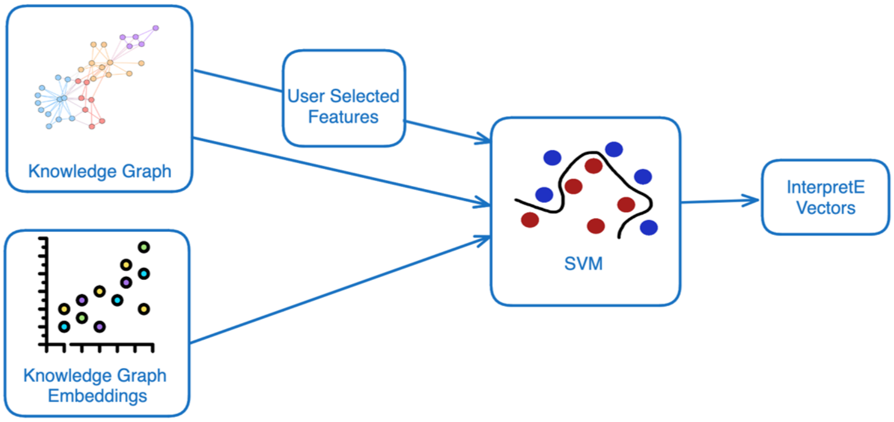


Fig. 1. Overview of *InterpretE*

human-understandable spaces. Conceptual spaces, introduced by Peter Gardenfors [10], represent concepts through cognitively meaningful features known as *quality dimensions*. These dimensions are typically learned from human judgments and serve as an intermediary representation layer between neural and symbolic representations. While promising for the advancement of explainable AI, this approach has not been extended to more complex datasets such as KGs and their representations. Our proposed approach is a first step towards identifying similar interpretable dimensions for KGEMs and deriving vector spaces that are human-understandable in terms of the underlying features of the KG entities.

3 *InterpretE*

In this section, we present the proposed *InterpretE* approach that aligns the vector representations with the entity features by the manipulation of the vector spaces. Figure 1 provides a simplified view of the proposed approach. Essentially, n-dimensional entity vectors from a given pre-trained KGEM serve as the input, along with a set of d features for these entities that are desired to be represented in the vector space (these can be task driven, e.g. separating players from politicians). Further details of the approach are provided below. An SVM model is trained on the vectors, guided by the features, to produce d-dimensional *InterpretE* vectors where the dimensions correspond to the entity features. Moreover, entities that are similar in terms of the specified features are clustered together. Further details of the approach are provided below.

Feature Selection. The *InterpretE* approach is centered around the representation of the desired features of the entities in the vector space. We designed several

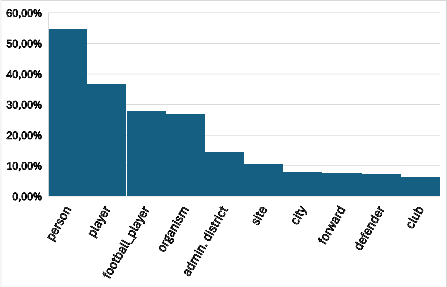


Fig. 2. Top 10 most represented entity types in Yago3-10

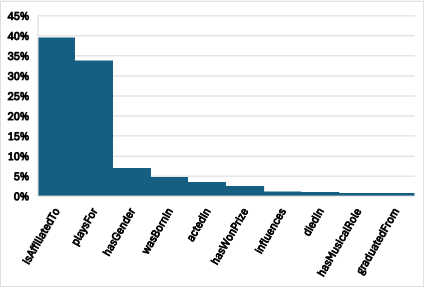


Fig. 3. Top 10 most represented relations for *person* entities in Yago3-10

experiments with different features to test the approach[4]. Feature selection was crucial as it guided experiment design.

As a first step, entities within the KG were categorized by their ontological classes using *WordNet* types such as persons, organizations, and locations. For each entity type, the most representative relations were selected and their values were categorized based on their distribution in KG triples. These categories served as the entity features that dictate the dimensions in the *InterpretE* vector spaces. An overview of the dataset analysis in terms of the most representative entity types for the Yago3-10 dataset is shown in Figure 2. Furthermore, the most significant relations for the person entities are shown in Figure 3.

For designing the experiments, different levels of abstraction were considered for features. For example, for *person* entities, the relation *'wasBornIn'* (e.g., *wasBornIn* Paris) was found significant. One experiment mapped locations from specific cities to their corresponding countries (e.g., France), while another grouped the cities to their respective continents (e.g., Europe), allowing evaluation across different abstraction levels. The different experiments will be presented and discussed in Section 4.

This adaptable process was primarily driven by the availability of sufficient data points for the features within the KG. Once features were defined, entities were labeled with binary values indicating the presence or absence of each feature. This labeled data was subsequently used for SVM training in the next phase.

Derivation of Interpretable Vectors. After identifying the features for different types of entities, we trained Support Vector Machine (SVM) classifiers on each feature using a training set, following a similar methodology to that used by Derrac et al. [7]. Although we have access to the ground truth for each vector, we chose to use SVM classifiers instead of directly converting the ground truth into binary vectors. This approach allowed us to retain more detailed information

[4] Note that the attributes of the KG entities could not be considered as features since most KGEMs are not trained on them, hence such features cannot be derived from the original vectors.

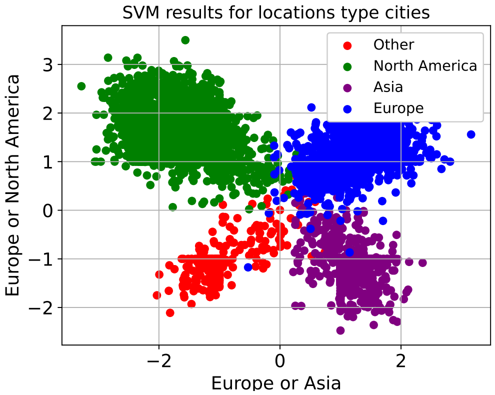


Fig. 4. Example 2D visualization of *InterpretE* vectors (*city* entities, *location* as features) in Yago3-10

encapsulated in the original embeddings, rather than reducing it to simple binary values.

To streamline the SVM training, we automated the process and defined a set of possible parameters for the SVM. Grid search and cross validation was performed in order to select the best estimator (with the Scikit-learn [19] library which uses LibSVM [6]). This methodology helped prevent overfitting and ensured a more generalized estimated hyperplane. To address class imbalance in the KG data, weights were assigned to entities based on their class distribution. The performance was evaluated using a test set comprising 20% of all entities (without any overlap with the entities in the train set). At the end of this process, new vectors were derived for each entity with each dimension corresponding to a specific feature and the sign indicating the associated feature.

4 Experiments

Datasets and Embeddings. To derive and categorize features for different entities in the KG, their type information was essential. As such, we leveraged KG datasets with associated ontologies, focusing on subsets of Yago (Yago3-10 [17]) and Freebase (FB15k-237) [22]. Additionally, we reused *Wordnet*-based entity type mappings from Jain et al. [14].

Following previous works [14, 12], several popular and benchmark KGEMs were considered for the experiments to analyse the scalability of the *InterpretE* approach across vector spaces generated with different methods, including ConvE [8], TransE [2], DistMult [26], Rescal [18] and Complex [24][5].

[5] The pretrained embeddings were obtained from `https://github.com/nitishajain/KGESemanticAnalysis`

Evaluation of InterpretE Vector Space. The derived *InterpretE* vector spaces are expected to cluster the vectors for the entities as per the selected features. An example for the 2D visualization of these clusters is shown in Figure 4, where the experiment centered around *city* entities and their *locations* as features (abstracted to continents). In order to evaluate these clusters and to guarantee a general space, the Cohen's kappa coefficient (κ score) was calculated for the test set (following [7]). This metric measures the agreement between two dependent categorical samples. The value ranges from -1 to 1, with a value closer to 1 indicating stronger agreement between the trained SVM and the ground truth on the testing set. The values of the mean κ score for the different experiments on Yago3-10 dataset are shown in Table 1. (The results for FB15k-237 are available in the appendix (Table 2)). Values close to 1 for this metric for most experiments indicates the promise of the approach.

Evaluation of Semantic Similarity. InterpretE vectors are dictated by the selected features for the entities that they represent, as such we evaluated the semantic similarity of the derived vectors (in terms of the features) to measure this desirable characteristic. We propose a simple metric *simtopk* to measure the similarity of entities' neighbors. For each entity, we analyze its neighborhood to estimate the similarity based on the corresponding feature used in the SVM experiment. The parameter k represents the number of neighbors considered. The score assigned to the original entity is calculated as the mean value of the similarities computed with these neighboring entities. This process is repeated for all entities, and the mean value of these scores is computed to serve as the final metric. The proposed *simtopk* metric can be formulated as:

$$\mathit{simtopk} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{k} \sum_{j \in N_i(k)} f(n_i, n_j) \right) \tag{1}$$

where: n : the number of total entities; k : the number of considered neighbours; $N_i(k)$: the k closest neighbours of the i-th entity, determined using a euclidean distance; $f(\cdot, \cdot)$: returns 1 if the two entities are similar in terms of features, 0 otherwise.

The values of this metric for k=10 for the *original* and the derived *InterpretE* embeddings for the different experiments and the various embedding models are shown in Table 1 for Yago3-10 (and Table 2 for FB15k-237 in appendix). The scores are better for *InterpretE* vectors as compared to the original pre-trained vectors (obtained from various KGEMs) across the board, indicating that similar entities are being represented by vectors that are closer in the new vector space, as desired.

An alternative way to evaluate the semantic similarity using large language models was also explored and is presented in the appendix (Appendix B).

Entity type and chosen features		ConvE	TransE	DistMult	Rescal	Complex
person : hasGender - wasBornIn (Europe)	κ score	.96	.93	.95	.96	.94
	original	.456	.496	.492	.507	.504
	InterpretE	**.54**	**.529**	**.538**	**.543**	**.539**
person : wasBornIn (Europe - Asia - North America)	κ score	.92	.84	.90	.94	.90
	original	.687	.8	.814	.871	.831
	InterpretE	**.987**	**.959**	**.983**	**.987**	**.979**
person : playsFor (UK - Germany - Italy - US)	κ score	.80	.80	.81	.80	.81
	original	.789	**.832**	.838	.828	.85
	InterpretE	**.917**	.716	**.913**	**.9**	**.942**
person : worksAt (university - educational_institution - organization)	κ score	.31	.13	.32	.31	.30
	original	.467	.413	.465	.461	.465
	InterpretE	**.868**	**.868**	**.853**	**.86**	**.807**
person : type (player - artist - politician - scientist - officeholder - writer)	κ score	.77	.75	.78	.78	.74
	original	.745	.772	.805	.794	.662
	InterpretE	**.953**	**.945**	**.958**	**.944**	**.938**
city : isLocatedIn (Europe - Asia - (North - South) America)	κ score	.94	.96	.96	.98	.98
	original	.899	.959	.949	.966	.972
	InterpretE	**.989**	**.993**	**.991**	**.996**	**.996**
organizations : location (US - UK - Canada - Japan - France - Australia)	κ score	.52	.53	.51	.58	.54
	original	.622	.694	.658	.703	.703
	InterpretE	**.904**	**.786**	**.912**	**.899**	**.897**
scientist : hasWonPrize	κ score	.96	.84	.97	.85	.98
	original	.539	.51	.575	.538	.578
	InterpretE	**.958**	**.934**	**.966**	**.926**	**.972**

Table 1. *simtop10* scores on the original and *InterpretE* vectors and κ scores on the test set for the experiments with Yago3-10 for the different KGEMs.

4.1 Discussion

The results from the designed experiments for each dataset demonstrate the potential of the proposed approach. However, there are several considerations for the experiment design that depend heavily on the data distributions and characteristics of the underlying KG data. For example, there is often class imbalance in entities concerning selected features (e.g., *hasGender* having more *male* representatives than *female*). These factors can impact the performance of the SVM classifier. Class-based weights have been applied to the data points to address this issue, but it remains a design challenge.

In some experiments, our method achieves a *simtopk* value very close to 1. This indicates that in the resulting space, similar entities are clustered together nearly perfectly. However, this level of clustering is not consistently observed across all experiments. The variability can be explained by the fact that other underlying features, not covered in the current experiment, could contribute to more accurately clustering similar entities. An analogy can be drawn with the well-known kernel trick used in SVMs, where an additional dimension (in our case, the consideration of a new feature) is introduced to better distinguish different labeled data (in this context, non-similar entities). Another challenge is the abstraction of features, especially if the underlying data is noisy and non-canonicalized (e.g., different labels for the same value such as 'UK' and 'United Kingdom'). Resolving these issues is crucial for creating useful feature categories. A potential limitation of this approach could be scalability. As the size of the knowledge graph (KG) increases, the time complexity of training the SVM also increases. The time complexity of SVM training is $O(n^2d)$, where n is the number of entities and d is the number of dimensions. Despite these challenges, *InterpretE* represents a significant step towards deriving interpretable vector spaces from KGEM vectors. It is flexible and applicable to any KGEM. We aim to further develop this approach to streamline the design and engineering process as well as improving its scalability across various datasets.

5 Conclusion and Future Work

This paper attempts to address the oft overlooked issue of lack of semantic interpretability in latent spaces generated by popular KG embedding techniques. The proposed *InterpretE* approach is shown to be capable of deriving interpretable spaces from existing KGEM vectors with human-understable dimensions that are based on the features in the underlying KG. Through the design and evaluation of different experiments, we have showcased the promise of the approach for encapsulating entity features in the vectors for different feature abstraction levels, customizable as per the dataset. By aiming to bridge the gap between entity representations and human-understandable features, *InterpretE* paves the way for enhanced understanding and utilization of KGEMs in various applications. Future research can further explore the implications of this approach and extend its applicability to broader contexts within the field of knowledge representation and reasoning.

Acknowledgements. This work was partly funded by the HE project MuseIT, which has been co-founded by the European Union under the Grant Agreement No 101061441. Views and opinions expressed are, however, those of the authors and do not necessarily reflect those of the European Union or European Research Executive Agency.

We are also thankful for access to King's Computational Research, Engineering and Technology Environment (CREATE). King's College London. (2024). Retrieved June 26, 2024, from `https://doi.org/10.18742/rnvf-m076`.

References

[1] Faisal Alshargi et al. "Concept2vec: Metrics for evaluating quality of embeddings for ontological concepts". In: *arXiv preprint arXiv:1803.04488* (2018).

[2] Antoine Bordes et al. “Translating embeddings for modeling multi-relational data”. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. NIPS’13. Lake Tahoe, Nevada: Curran Associates Inc., 2013, pp. 2787–2795.

[3] Armand Boschin et al. “Combining embeddings and rules for fact prediction”. In: *International Research School in Artificial Intelligence in Bergen*. 2022.

[4] Zied Bouraoui, Víctor Gutiérrez-Basulto, and Steven Schockaert. “Integrating Ontologies and Vector Space Embeddings Using Conceptual Spaces”. In: *International Research School in Artificial Intelligence in Bergen (AIB 2022)*. Ed. by Camille Bourgaux, Ana Ozaki, and Rafael Peñaloza. Vol. 99. Open Access Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 3:1–3:30. ISBN: 978-3-95977-228-0. DOI: `10.4230/OASIcs.AIB.2022.3`. URL: `https://drops.dagstuhl.de/entities/document/10.4230/OASIcs.AIB.2022.3`.

[5] Zied Bouraoui et al. “Modelling semantic categories using conceptual neighborhood”. In: Cited by: 8. 2020, pp. 7448–7455. URL: `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85100751171&partnerID=40&md5=b3889af3050ba94181fc4ff357a1fdb9`.

[6] Chih-Chung Chang and Chih-Jen Lin. “LIBSVM: A library for support vector machines”. In: *ACM Trans. Intell. Syst. Technol.* 2.3 (May 2011). ISSN: 2157-6904. DOI: `10.1145/1961189.1961199`. URL: `https://doi.org/10.1145/1961189.1961199`.

[7] Joaquín Derrac and Steven Schockaert. “Inducing semantic relations from conceptual spaces: A data-driven approach to plausible reasoning”. In: *Artificial Intelligence* 228 (2015), pp. 66–94. ISSN: 0004-3702. DOI: `https://doi.org/10.1016/j.artint.2015.07.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0004370215001034`.

[8] Tim Dettmers et al. *Convolutional 2D Knowledge Graph Embeddings*. 2018. arXiv: `1707.01476 [cs.LG]`.

[9] Mohamed H Gad-Elrab et al. “Excut: Explainable embedding-based clustering over knowledge graphs”. In: *International Semantic Web Conference*. Springer. 2020, pp. 218–237.

[10] Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. The MIT Press, Mar. 2000. ISBN: 9780262273558. DOI: `10.7551/mitpress/2076.001.0001`. URL: `https://doi.org/10.7551/mitpress/2076.001.0001`.

[11] Xiou Ge et al. *Knowledge Graph Embedding: An Overview*. 2023. arXiv: `2309.12501 [cs.AI]`.

[12] Nicolas Hubert et al. *Do Similar Entities have Similar Embeddings?* 2024. arXiv: `2312.10370 [cs.AI]`.

[13] Filip Ilievski et al. “A study of concept similarity in Wikidata”. In: *Semantic Web* (Jan. 2024), pp. 1–20. DOI: `10.3233/SW-233520`.

[14] Nitisha Jain et al. “Do Embeddings Actually Capture Knowledge Graph Semantics?” In: *The Semantic Web*. Ed. by Ruben Verborgh et al. Cham:

Springer International Publishing, 2021, pp. 143–159. ISBN: 978-3-030-77385-4.

[15] Albert Q. Jiang et al. *Mistral 7B*. 2023. arXiv: `2310.06825 [cs.CL]`.

[16] Jan-Christoph Kalo, Philipp Ehler, and Wolf-Tilo Balke. "Knowledge graph consolidation by unifying synonymous relationships". In: *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18*. Springer. 2019, pp. 276–292.

[17] Farzaneh Mahdisoltani, Joanna Asia Biega, and Fabian M. Suchanek. "YAGO3: A Knowledge Base from Multilingual Wikipedias". In: *Conference on Innovative Data Systems Research*. 2015. URL: `https://api.semanticscholar.org/CorpusID:6611164`.

[18] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. "A Three-Way Model for Collective Learning on Multi-Relational Data". In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. Ed. by Lise Getoor and Tobias Scheffer. ICML '11. Bellevue, Washington, USA: ACM, June 2011, pp. 809–816. ISBN: 978-1-4503-0619-5.

[19] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[20] Adi Simhi and Shaul Markovitch. *Interpreting Embedding Spaces by Conceptualization*. 2023. arXiv: `2209.00445 [cs.CL]`.

[21] Zequn Sun et al. "A benchmarking study of embedding-based entity alignment for knowledge graphs". In: *arXiv preprint arXiv:2003.07743* (2020).

[22] Kristina Toutanova and Danqi Chen. "Observed versus latent features for knowledge base and text inference". In: *Workshop on Continuous Vector Space Models and their Compositionality*. 2015. URL: `https://api.semanticscholar.org/CorpusID:5378837`.

[23] Hugo Touvron et al. "LLaMA: Open and Efficient Foundation Language Models". In: *ArXiv* abs/2302.13971 (2023). URL: `https://api.semanticscholar.org/CorpusID:257219404`.

[24] Théo Trouillon et al. "Complex Embeddings for Simple Link Prediction". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 2071–2080. URL: `https://proceedings.mlr.press/v48/trouillon16.html`.

[25] Quan Wang et al. "Knowledge graph embedding: A survey of approaches and applications". In: *IEEE transactions on knowledge and data engineering* 29.12 (2017), pp. 2724–2743.

[26] Bishan Yang et al. *Embedding Entities and Relations for Learning and Inference in Knowledge Bases*. 2015. arXiv: `1412.6575 [cs.CL]`.

[27] Xinqi Zhu, Chang Xu, and Dacheng Tao. *Where and What? Examining Interpretable Disentangled Representations*. 2021. arXiv: `2104.05622 [cs.CV]`. URL: `https://arxiv.org/abs/2104.05622`.

Appendix A Statistics and Results for FB15K-237

To study the scalability of the *InterpretE* method across different KGs, we also experimented with the FB15K-237 dataset, following the same design methodology as with Yago3-10. In this section, we present the statistics and the results obtained on this dataset.

Similar to Yago3-10, we conducted a statistical analysis to select features. Figure 5 illustrates the most represented entity types in this dataset. For each type considered, we identified the most represented relations, detailed in Figure 6 for *film* entities as an example.

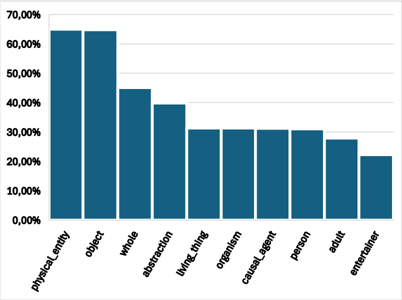


Fig. 5. Top 10 most represented entity types in FB15K-237

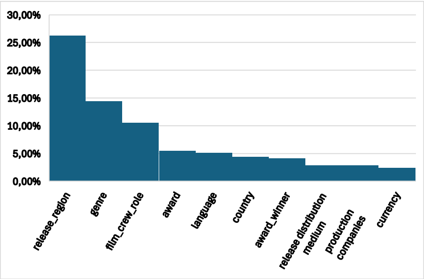


Fig. 6. Top 10 most represented relations for *film* entities in FB15K-237

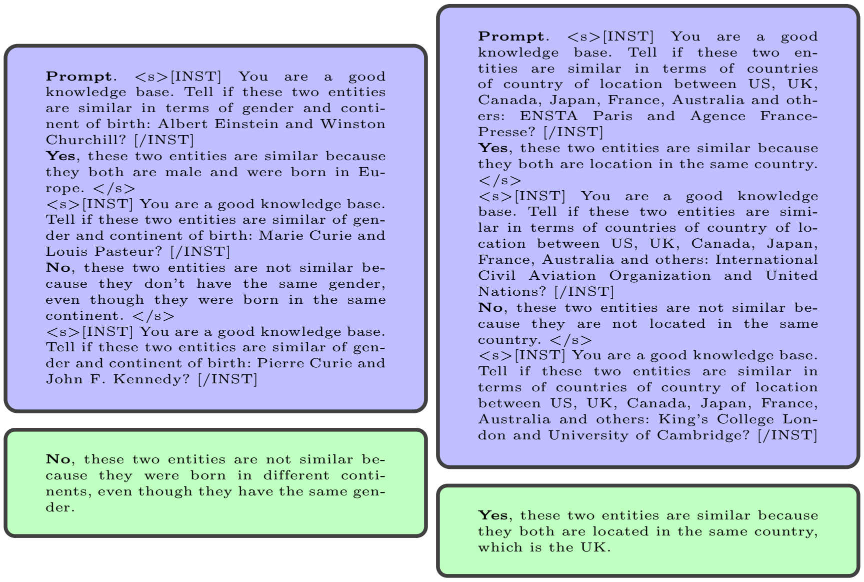
Following the training phase, we evaluated the semantic similarity using the *simtopk* metric and obtained results as presented in Table 2. Similar to our findings with Yago3-10, we observed enhanced semantic similarity with FB15K-237. This improvement is evidenced by the higher *simtopk* value in the final space compared to the original space.

Entity type and chosen features		ConvE	TransE	DistMult	Rescal	Complex
person : gender - nationality (USA - England - UK - India - Canada)	κ score	.84	.73	.83	.88	.84
	original	.587	.524	.575	.563	.563
	InterpretE	**.952**	**.918**	**.936**	**.956**	**.932**
organizations : locations (USA - UK - Japan - Canada - Germany)	κ score	.78	.70	.75	.58	.79
	original	.766	.738	.758	.731	.768
	InterpretE	**.951**	**.947**	**.958**	**.959**	**.96**
film_release_region (USA - Sweden - France - Spain - Finland)	κ score	.71	.69	.71	.66	.71
	original	.705	.66	.661	.621	.661
	InterpretE	**.876**	**.866**	**.903**	**.907**	**.892**
film genre (drama - comedy - romance - thriller - action)	κ score	.68	.65	.71	.72	.70
	original	.212	.217	.215	.217	.213
	InterpretE	**.732**	**.719**	**.805**	**.78**	**.753**

Table 2. *simtop10* scores on the original and *InterpretE* vectors and κ scores on the test set for the experiments with FB15K-237 for different KGEMs.

Appendix B Semantic Similarity Evaluation with LLMs

We also explored an alternative way to the *simtopk* metric using a large language model (LLM) in a limited experiment. We attempted this approach with few-shot prompting using Llama3-70B [23]. Additionally, we experimented with a RAG pipeline using the entire initial knowledge graph with Mistral7B [15]



Prompt. <s>[INST] You are a good knowledge base. Tell if these two entities are similar in terms of gender and continent of birth: Albert Einstein and Winston Churchill? [/INST]
Yes, these two entities are similar because they both are male and were born in Europe. </s>
<s>[INST] You are a good knowledge base. Tell if these two entities are similar of gender and continent of birth: Marie Curie and Louis Pasteur? [/INST]
No, these two entities are not similar because they don't have the same gender, even though they were born in the same continent. </s>
<s>[INST] You are a good knowledge base. Tell if these two entities are similar of gender and continent of birth: Pierre Curie and John F. Kennedy? [/INST]

No, these two entities are not similar because they were born in different continents, even though they have the same gender.

Prompt. <s>[INST] You are a good knowledge base. Tell if these two entities are similar in terms of countries of country of location between US, UK, Canada, Japan, France, Australia and others: ENSTA Paris and Agence France-Presse? [/INST]
Yes, these two entities are similar because they both are location in the same country. </s>
<s>[INST] You are a good knowledge base. Tell if these two entities are similar in terms of countries of country of location between US, UK, Canada, Japan, France, Australia and others: International Civil Aviation Organization and United Nations? [/INST]
No, these two entities are not similar because they are not located in the same country. </s>
<s>[INST] You are a good knowledge base. Tell if these two entities are similar in terms of countries of country of location between US, UK, Canada, Japan, France, Australia and others: King's College London and University of Cambridge? [/INST]

Yes, these two entities are similar because they both are located in the same country, which is the UK.

Fig. 7. Partial example of few-shot prompts with Llama 3 70B using HuggingChat

and LlamaIndex. However, the results were not consistently convincing, and the model sometimes contradicted itself.

In our prompt to the LLM, we provided two examples: one positive and one negative, randomly chosen from all possible entities. We also specified the type of similarity we were evaluating, as it depended on the selected feature for a given experiment. This method allows us to assess our approach by examining how similar the neighborhood of a given entity is to the entity itself. This approach needs to be applied to all entities to obtain a global evaluation metric, which we plan to explore in future work.