Investigating Latent Representations under Structured Sampling and Constraints

Antrea Christou[1], David Castro[1] and Cogan Shimizu[1,*,†]

[1]Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435, United States of America

Abstract

The explainability and interpretability of learned latent spaces remain important and active avenues of investigation. In this work, we examine the effect of various materialization techniques that we hypothesize rigorously and consistently impact learned representations of knowledge graphs. In particular, we examine the effect of materializing the transitive closure for various relations over numeric literals-as-objects. We also consider the impact of subsumptive windowing techniques. We test these theories through a series of controlled experiments over synthetic knowledge graphs (SKGs) representing persons and their ages. Our results demonstrate consistent ordering in the latent space, and the capacity to learn mappings across different monotonic number distributions.

Keywords

Knowledge Graphs, Synthetic Knowledge Graphs, Knowledge Graph Embeddings, MuRE, Monotonic Relationships

CEUR-WS.org/Vol-4313/kgnesy-paper19.pdf

1. Introduction & Motivation

Knowledge Graph Embeddings (KGEs) facilitate the exploration of data and are a convenient learned representation for a variety of tasks (e.g., link prediction [1] or entity clustering [2]). However, there is a plethora of models, often grouped by machine learning task or differentiated based on target graph characteristics. What information (or semantics) is preserved from the structure of our data into the vector itself still remains unclear [3]. This work is motivated by the deeper investigation between the "actual" semantics and the "learned" results the KGEs produce. We examine techniques for enforcing distinct structures[1] in a learned latent space. Generally speaking, we search for a contrived scenario which is amenable to this, but is still representative of a real-world KG, and as such produce recommendations to impact KGE outcomes. More specifically, we explore how numerical may be preserved in the embedding, so as to have a better understanding on how this "learning" is effectuated.

To isolate the behavior of knowledge graph embeddings on numerical information, we construct Synthetic Knowledge Graphs (SKGs) in which age is represented as a numeric literal. Age offers a continuous, monotonic attribute with known ground-truth ordering, allowing us to systematically evaluate both structure preservation and numeric inference. Consistent removal of age relations further enables simulation of different supervision levels while maintaining a controlled experimental setting.

We build upon the state of the art in understanding and evaluating KGEs [4, 5, 6, 7, 8]. This work aims to address the following research questions (RQs), at a broad level:

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

*Corresponding author.

christou.2@wright.edu (A. Christou); castro.31@wright.edu (D. Castro); cogan.shimizu@wright.edu (C. Shimizu)

0009-0003-0428-2439 (A. Christou); 0009-0006-8593-4188 (D. Castro); 0000-0003-4283-8701 (C. Shimizu)

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

[1]For lack of a better term, we use *structure* to refer to 2D-shapes (or 3D-clouds) of vector terminals, which we posit can then be interpreted.

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-05

RQ1. To what extent can the order of a monotonic sequence of numeric literals be preserved in the latent space?
RQ2. How does the materialization of a different ontology commitments (i.e., hierarchical window construction and transitive closure over relations) influence the latent space projections?
RQ3. How accurately can numeric values be inferred from the latent space, at various levels of supervision restriction?

To systematically investigate these RQs, we have formulated corresponding hypotheses, providing the basis for our empirical evaluation.

H1. Materialization of monotonic numeric relations reduces ordering inconsistencies between the ground-truth numeric values and their corresponding latent representations.
H2. Encoding transitive relations before training a KG with a specific KGE model results in stable numeric representations in the latent space.
H3. Numeric information encoded within the latent space remains recoverable under partial supervision.

We test these hypotheses via a series of experiments that investigate how knowledge graph construction and materialization strategies influence the information preserved within learned latent representations. These ultimately result in a set of concrete deliverables that make these experiments measurable and reproducible, as follows:

D1. Method: A Synthetic Knowledge Graph construction (SKG) framework; Modeling numeric literals (ages) including relational removal patterns that can be tweaked. This is in order to simulate different supervision levels. This includes materialization strategies for numeric values, implementation of window-based *lessThan* relation hierarchies and transitive order relations (ages between people).
D2. Evaluation: A systematic framework for evaluating monotonicity preservation, inference performance, statistical significance (paired t-tests and Wilcoxon signed-rank tests), and scalability (experiments across varying dataset sizes).
D3. Code and Reproducibility: All of our experiments, instructions, results and figures are gathered and displayed in our repository https://github.com/kastle-lab/KGE-Neurosymbolic-AI.

The rest of the paper is organized as follows. Section 2 gives information regarding aspects of related work that have influenced this work, with statements on how we differ. Section 3 presents the methodology, including the construction of our SKGs and how we encoded our numeric information. Section 4 defines the metrics we used to investigate our method. Section 5 state the results we obtained along with figures showcasing visualization of those results. Section 6 summarizes the findings and states future directions.

2. Related Work

This section briefly reviews related or parallel efforts to this work, including surveys on KGE models that encode literals, neurosymbolic and logic-based approaches, and other latent space analysis.

KGE Models and Structural Design. Gesese et al. [9] survey literal-aware KGE models, emphasizing methods that incorporate textual and numerical literals to improve link prediction, while Christou et al. [7, 4] demonstrate that graph modeling choices, such as contextual nodes, reification, and shortcuts, significantly influence embedding performance. Gelal et al. [10] further show that increased model complexity does not necessarily improve downstream application performance, highlighting the importance of both embedding architecture and graph design.

Our work investigates how graph construction, materialization, and hierarchical numerical literals influence the structural properties encoded in the latent space, extending the evaluation beyond predictive performance to include numerical ordering, consistency, and interpretability, shifting the

solely focus from which models perform better to what information is actually encoded in the latent space.

Neurosymbolic Reasoning and Latent Space Interpretability. Research conducted by Ben Chaabene et al [11] presents a neurosymbolic framework that enhances the explainability and interpretability of model predictions by combining KGEs with logical reasoning. Along with that, Munda et al. [12] explore the use of vector space transformations to extract structured knowledge graphs from the latent representations of neural language models. Another approach by Hubert et al [13], examines the prospective benefits of using domain and range constraints in negative sampling in order to optimize link prediction results.

Our research takes a complementary approach, introducing symbolic structure at the graph construction level and examining how the learned embedding space reflects these created constraints.

3. Methodology

This section reports the construction method of our two SKG variants: SKG KG_s and SKG KG_p. This incorporates details on how we created the KGs along with visualizations of their entities and relationships.

3.1. Numeric Value Encoding

Understanding how structured and symbolic information is reflected in latent spaces is a common objective of our work, but instead of using pre-trained KGS or any commonly used benchmarks, we investigate how controlled graph construction and materialization strategies affect retention of such structure during the training process. We construct two synthetic KGs: SKG KG_s and SKG KG_p where we encoded continuous numeric information in order to properly investigate the preservation of that after training.

3.1.1. Creating SKG KG_s

We define KG_s as follows. First, we define any KG to have the form $KG = (N, R, T)$, where N is a set of nodes, R is a set of relations, and T is a set of triples over $N \times R \times N$. For KG_s, $R = \{\mathsf{lessThan}, \mathsf{partOf}\}$, each $n_i \in N$ is augmented with a value $x_i \sim U[0, 1]$, and define a function f such that $f(n_i) = x_i$. We continue by defining an ordering over N from f. where lessThan triples are therefore built for each value-augmented node $n_i \in N$ pairwise to every other node that satisfy the lessThan condition. That is,

$$\{(n_a, n_b) \in N^2 | f(n_a) < f(n_b)\} \implies (n_a, \mathsf{lessThan}, n_b) \in KG_s. \tag{1}$$

More formally, " $\implies$ " demonstrates how to transform the ordering into a KG triple, and "$\in$" simply means the triple is to be inserted into the KG. For example, if n_1 is randomly assigned a value of 0.5 and n_2 is randomly assigned a value of 0.2, then a triple is generated consisting of:

$$(f(n_2) < f(n_1)) \implies (n_2, \mathsf{LessThan}, n_1) \in KG_s \tag{2}$$

Moreover, the set of value-assigned nodes was divided recursively at depth d based on their assigned values, and each assigned-value node was associated to a "window" node to construct connections. In other words, the entire set of assigned-value nodes are each related to the window_root node; the lower half of nodes are each related to window_root_left, and conversely, the right half are related to window_root_right after ordering nodes according to their assigned values. Let n_median denote the

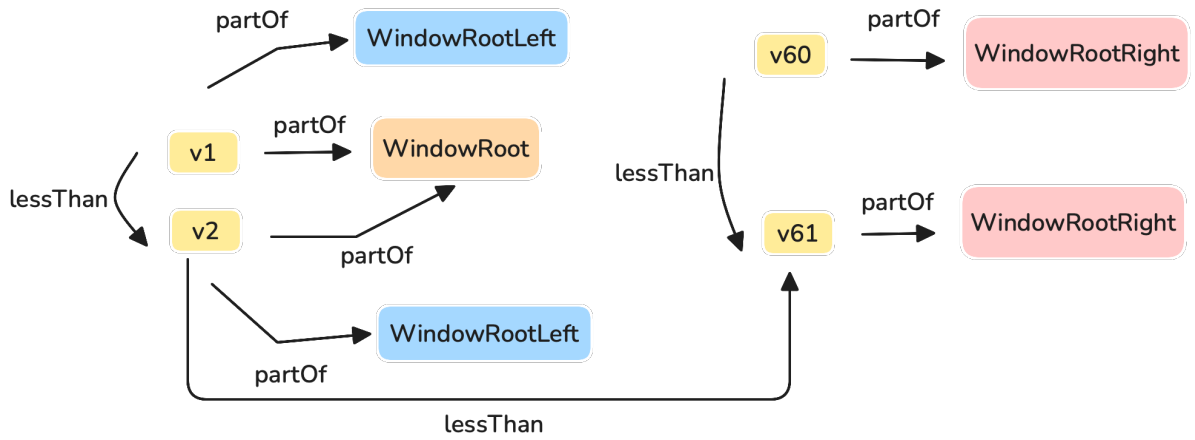


Figure 1: Example of the window construction for SKGs

node whose assigned value is the median of the ordered node values within the current partition:

$$(n_a < n_{median} < n_b) \implies (n_a,\ \text{partOf},\ w_{left}) \in KG_s \wedge (N_b,\ \text{partOf},\ w_{right}) \in KG_s \tag{3}$$

Every half is further split in the same order, leading to 2^d sub-windows + 1 root-window. Once all LessThan and partOf relationships have been settled, the assigned values are set aside while keeping the structure from lessThan and partOf relations as they are. Essentially, we are strictly defining their position relative to each other and ensuring this is preserved in the embedding itself.

3.1.2. Creating SKG KG_p

To further examine the potential structures in the latent space, we embed an SKG with a similar structure to KG_s. This new SKG, which we call KG_p follows can be interpreted as an example. Previously, KG_s is simply an ordering over a uniform distribution of real values on the unit interval. However, now, we can consider that this sampling was instead a normalization over some meaningful data. In this case, it is easy to understand these as *ages*. KG_p can be considered to be a "supergraph" of KG_s. Specifically, let $KG_p = (N_p, R_p, T_p)$ and similarly for KG_s. Then, $N_s \subset N_p$ and $T_s \subset T_p$. Additionally, we add two new relations to R: youngerThan and hasAge. Of course, there are not always as exactly as many things (in this case, people) as there are ages, so we also parameterize $KG_{p,n}$, where n is the size of the population of entities with an age. To simplify our construction from now, instead of treating data as a real value, we define as an integer in $[0, 100]$ These ages follow the same construction as above.

$$(f(n_a) < f(n_b)) \implies (n_a, \text{lessThan}, n_b) \in KG_p \tag{4}$$

Then, we generate n nodes (i.e., n people) that have an age. Rather than the "augmentation method" used in for KG_s, these are defined explicitly. For each person, they are assigned a random age, and the following triple is added.

$$(P_i) \implies (P_i, \text{hasAge}, n_j) \tag{5}$$

Then, the person nodes are ordered based on their explicitly associated age triple:

$$(P_a < P_b) \implies (P_a, \text{youngerThan}, P_b) \in KG_p \tag{6}$$

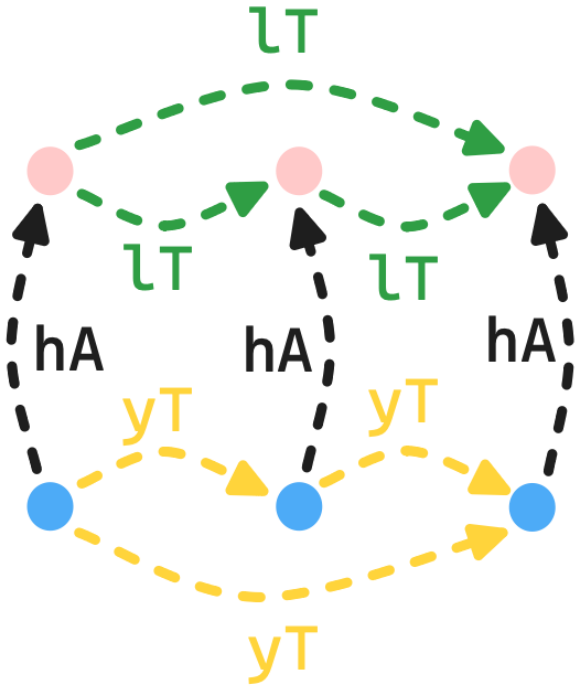


Figure 2: Construction of $SKG_p, 3$ indicating the encoding of the ages along with lessThan and youngerThan relationships. The blue entities represent the people and the pink ones represent the ages.

4. Evaluation

To formally assess our modeling choices, answer our RQs, and confirm our hypotheses we took the followed steps for each SKG and chose a variety of suitable metrics that measure how "well" do the projected values lie in the latent space.

4.1. Experimental Setup

This section describes the experimental procedures and evaluation criteria used to assess whether numeric information is preserved and recoverable within the learned latent representations. We evaluate the embeddings from both SKGs through complementary analyses: structural preservation, which examines whether the original numeric ordering is retained, and structural coherence, which evaluates the ability to infer missing numeric information under varying levels of supervision.

4.1.1. Embedding KG_s and KG_p

We chose MuRE [14] as the most appropriate KGE model since it incorporates relation-specific scaling and translation operations. While we tried models such as TransE [15] and TransR [16], the results were not as prominent. Those results reside in our repository for comparison. MuRE adapts better in more complex hierarchical structures like the age relationships between different people.

For example, for a query such as:

(person42, hasAge, ?)

MuRE transforms the embedding of the head entity using the learned representation of the relation:

$$R \odot h + r \tag{7}$$

where:

1. ‘h‘ is the embedding of the head entity
2. ‘R‘ is a relation-specific scaling vector
3. ‘r‘ is a relation-specific translation vector
4. ‘$\odot$‘ represents element-by-element multiplication

4.1.2. Structural Preservation Metrics

Structural Preservation Metrics aim to measure the extend of which hierarchical and numeric structure is preserved within the embedding space. Assigned-value embeddings from *SKG* KG_s are analyzed using both learned nonlinear projections and linear dimensionality reduction techniques. Specifically, embeddings are projected onto an MLP-fitted curve as well as a one-dimensional PCA projection, enabling comparison between the induced latent ordering and the ground-truth ordering derived from assigned values.

- *PCA Projection:* Evaluates whether the embedding space naturally contains a linear direction corresponding to the ordering of the encoded numeric values. The first principal component is used as a one-dimensional projection of the embeddings, where higher rank correlation with the original values indicates stronger linear preservation of numeric structure.
- *MLP Projection:* Evaluates whether a nonlinear function can recover the ordering of the encoded numeric values from the learned embeddings. The MLP learns a mapping from the embedding space to the original values, where stronger correlation between the predicted values and ground-truth ordering indicates that the numeric structure is recoverable through nonlinear transformations.

4.1.3. Measures of Structural Coherence

Experiments are conducted across multiple graph variations with differing levels of supervision and materialization. These include fully supervised graphs as well as progressively sparse configurations where subsets of `hasAge` relations from *SKG* KG_p are removed prior to training. Evaluations are performed both over all entities and specifically over removed relations to isolate true inference behavior.

Our evaluation was conducted across three variants of KG_p: 100, 200, 500 P nodes: KG_{p100}, KG_{p200}, and KG_{p500}. These variations are essential to ensure that our experiments are reliable by reducing dependence on a single experimental configuration.

We calculated both Spearman rank correlation (rho) [17] and Kendall rank correlation (tau) [18], to determine the numerical agreement between the resulting embedding ordering and the underlying truth ordering. While Kendall's tau assesses pairwise ordering consistency using compatible and dissimilar pairs, Spearman's (rho) evaluates the degree of consistency among a pair of ordered sequences.

We have also formulated a *Prediction Task* as an effort of assessing how well is the age information in the latent space when we have partial supervision. Essentially, we formulated said task as predicting the missing age relationships for the *Person* nodes that are in the form :

$$(person_i,\ hasAge,\ v_j) \tag{8}$$

Where v_j represents the node corresponding to age j. During modified runs, some of these relations are removed before training. The removal patterns were as followed:

- `0%`: no relations removed
- `15%`: 15% of relations removed
- `30%`: 30% of relations removed
- `45%`: 45% of relations removed

- 60%: 60% of relations removed
- 75%: 75% of relations removed

The evaluation was performed under progressively increasing levels of missing supervision, ranging from complete information to the removal of 75% of age relations. Therefore the evaluation was performed under two scopes: across all relations and specifically on the missing relations only, with the missing-only condition isolating the true inference capability of the models. We also monitored the difference between "*Window*" vs "*Without Window* results. In this paper, we showcase the *Window* results across all relations instead of missing-only to provide a consistent overview of inference performance across different supervision levels. The rest reside in our repository for completeness.

In order to determine whether learned mappings or geometric transformations are better at preserving age structure, we compared two prediction approaches. Using the *MuRE Query-Point Method*, The transformed query point is:

$$q_i = p_i \odot R + r \tag{9}$$

While the predicted age is:

$$\hat{y}_i = \underset{v_j}{\operatorname{argmin}} \|q_i - v_j\| \tag{10}$$

This evaluates whether the relational geometry alone encodes numeric structure. On the other hand, the learned linear mapping method predicts age through:

$$\hat{y}_i = f(q_i) \tag{11}$$

Where f is a ridge regression model trained on available *hasAge* relations.

Moreover, we measured *Prediction Error* using the absolute error between the actual and predicted age values, which is defined as:

$$|y_i - \hat{y}_i| \tag{12}$$

Afterwards we summarized the overall performance with calculating the *Mean Absolute Error*:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{13}$$

Where lower values indicate better performance.

Along with that, we provide the standard deviation (SD) of the absolute prediction errors to quantify the variability of each method's predictions around its mean error. For each prediction approach, the reported SD is calculated as the sample standard deviation of the individual absolute errors:

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (|y_i - \hat{y}_i| - \overline{e})^2} \tag{14}$$

where $|y_i - \hat{y}_i|$ represents the absolute prediction error for instance i, and $\overline{e}$ is the mean absolute error.

From those we computed the *Delta MAE* to quantify the improvement of the learned regression method relative to the query-point method. It is computed as the difference between the mean absolute errors of the two approaches:

$$\Delta\text{MAE} = \text{MAE}_{QP} - \text{MAE}_{LR} \tag{15}$$

where MAE_{QP} represents the error of the query-point method and MAE_{LR} represents the error of the learned regression method. Positive values indicate that the learned regression method achieves lower prediction error and therefore outperforms the query-point approach, while negative values

indicate that the query-point method performs better. Larger positive values correspond to greater improvements obtained by the learned regression method.

We continued with applied statistical tests as well to detect consistency and any performance differences across methods. One of those is the *Paired t-test* [19], used to assess whether the mean difference between paired errors is significantly different from zero, computed as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{16}$$

Followed by the *Wilcoxon Signed-Rank Test*, which evaluates consistent differences without assuming normality.

We have additionally included an effect size to measure the degree of performance differences between the two approaches, independent of sample size.

$$d_z = \frac{\bar{d}}{s_d} \tag{17}$$

Where $\bar{d}$ is the mean of the paired differences and s_d is the standard deviation. his complements the statistically significant tests by providing a consistent measure of how much one approach performs better than the other. We interpret the values as follows:

$\lvert d_z \rvert$	Interpretation
0.2	Small Effect
0.5	Medium Effect
0.8	Large Effect

Last but not least, we computed *Confidence Intervals* to calculate the level of uncertainty in both methods measured performance differences.

$$[CI_{low},\ CI_{high}] \tag{18}$$

Where intervals excluding zero indicate stable differences.

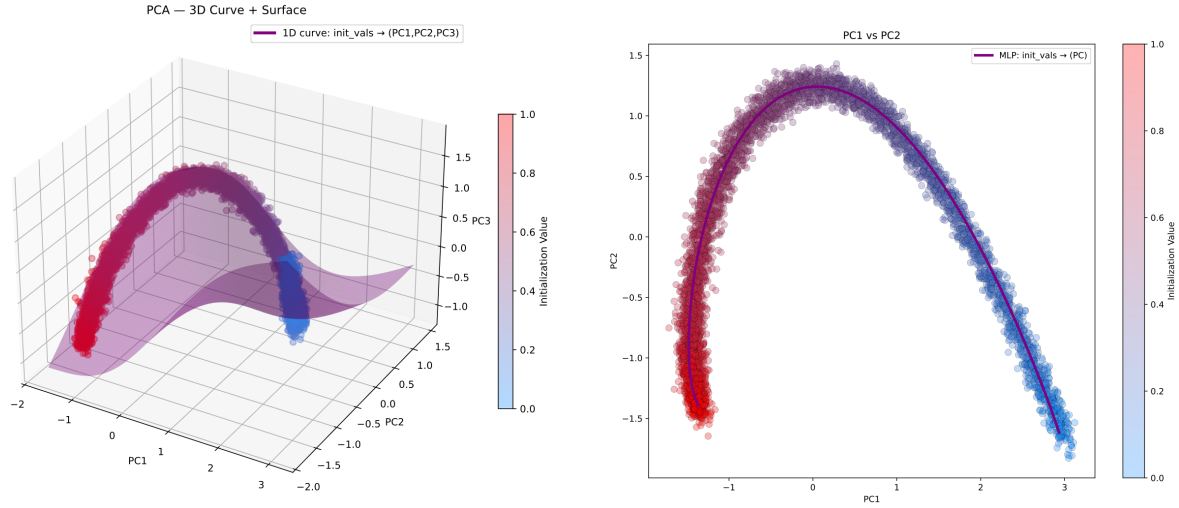
5. Results & Discussion

The embedding evaluation results, including PCA-based projections for SKG KG_s, quantitative inference metrics contrasting Query-Point and Learned Regression techniques under various experimental circumstances for SKG KG_p, are shown in this section along with interpretation and discussion on them. Mappings of results and RQs/Hs are also stated in this section.

5.1. Structure Preservation Results with respect to SKG KG_s

We first take a look at PCA component combinations in order to determine how well the learned embeddings align with the underlying ordered structure. Thus, we evaluate RQ 1 and RQ 2 defined in Section 1.

3D PCA Figure 3a demonstrates how the first three principal components are aligned with the original random initialization values (0 to 1), which are represented by the color gradient from blue to red, and a clear, sequential 1D path (the purple curve). The fact that the data points exhibit a clear, non-linear direction across the PC space, indicates that the structural relations; more specifically both the lessThan and hierarchical partOf constraints were successful in guiding the high-dimensional embeddings to maintain the original initialized underlying numerical form.



(a) 3D PCA projection.

(b) PC1 versus PC2 projection.

Figure 3: Principal Component Analysis (PCA) visualizations of the learned embeddings. (a) shows the three-dimensional projection onto the first three principal components, while (b) presents the two-dimensional projection onto the first and second principal components.

2D PCA Figure 3b displays pairs of principal components, offering different perspectives on the embedding space. The projections display the geometric structure of the embeddings, with nonlinear shapes and continuous arcs discernible throughout the point ranges. These shapes are closely followed by the corresponding MLP fit, which demonstrate the projection's general shape without entirely depending on particular embedding positions. They indicate that the embedding model has learned a structured latent space that preserves the ordered relationships present in SKG KG_s.

The above, support both H 1 and H 2 qualitatively, by showcasing the impact of materializing the monotonic relationships using hierarchical windows and theretofore maintaining a consistent geometric organization after training. Figure 4a showcases the plot of the first Principal Component (PC1) against the ground truth initial value across the population sizes. We can see a nearly linear monotonic correlation where lower PC1 values correspond to the red larger age groups and the larger PC1 values correspond to the blue lower age groups. This representation demonstrates that the underlying numerical encoding is indeed reflecting the result that is dominant by the ordering of the numbers rather than the windowing aspect.

To complement the PCA visualizations, we quantitatively evaluated whether the learned latent ordering agrees with the ground-truth ordering by reporting high Spearman and Kendall correlation numbers ρ= 0.975 and τ= 0.868 respectively as seen in Figure 4b. We also notice that the Multi-Layer Perceptron (MLP) line tracks and captures smoothly the age progression . This further confirms that the ordering induced by the embeddings closely matches the original numeric ordering.

5.2. Structural Coherence results with respect to SKG KG_p

The quantitative results for structural consistency are reported in Table 1, which summarizes QP and LR inference performance across varying population sizes and removal percentages, followed by visual analyses of the corresponding embedding structures. Metrics are introduced in Section 4.1.3 along with how are "good" results defined. Thus we evaluate RQ 3 defined in Section 1.

The numbers shown in Table 1 showcase that the Learned Regression (LR) Approach outperforms the Query-Point (QP) one across all population sized and removal percentages. At complete supervision i.e. 0% Removal across all population sizes reduced the MAE achieved with LR. As the removal of relation progressed, so did the error for LR. The opposite happened with QP. The Δ was the lowest

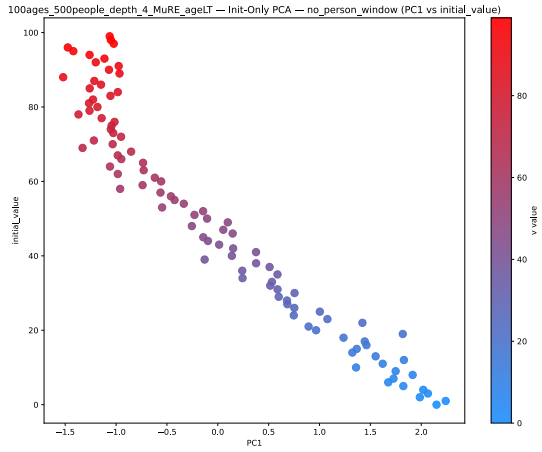
Population Size	Removal (%)	QP MAE/SD	LR MAE/SD	Δ MAE (QP–LR)	95% CI	Paired t p	Wilcoxon p	Cohen's d_z
500	0%	1.632, 3.333	0.929, 0.726	0.703	[0.418, 1.011]	<0.0001	0.0029	0.207
500	15%	2.698, 4.910	0.871, 0.801	1.827	[1.416, 2.268]	<0.0001	0.0096	0.378
500	30%	3.530, 5.101	1.355, 1.768	2.175	[1.758, 2.617]	<0.0001	<0.0001	0.443
500	45%	4.982, 6.600	1.603, 2.132	3.379	[2.811, 3.960]	<0.0001	<0.0001	0.521
500	60%	5.160, 5.822	1.896, 2.268	3.264	[2.763, 3.796]	<0.0001	<0.0001	0.545
500	75%	8.754, 8.203	1.736, 1.810	7.018	[6.322, 7.751]	<0.0001	<0.0001	0.871
200	0%	1.290, 3.740	0.471, 0.518	0.819	[0.328, 1.362]	0.0022	<0.0001	0.219
200	15%	1.920, 4.139	0.835, 1.619	1.085	[0.608, 1.597]	<0.0001	0.0124	0.313
200	30%	2.905, 5.074	1.313, 2.257	1.592	[0.973, 2.271]	<0.0001	0.9854	0.337
200	45%	4.480, 5.545	1.834, 2.741	2.646	[1.924, 3.409]	<0.0001	<0.0001	0.496
200	60%	5.835, 6.914	2.022, 2.531	3.813	[2.939, 4.735]	<0.0001	<0.0001	0.580
200	75%	8.730, 8.397	2.446, 2.492	6.284	[5.133, 7.484]	<0.0001	<0.0001	0.739
100	0%	2.790, 5.470	0.209, 0.374	2.581	[1.592, 3.646]	<0.0001	0.7129	0.478
100	15%	3.490, 5.663	0.765, 1.844	2.725	[1.751, 3.787]	<0.0001	0.0077	0.519
100	30%	6.200, 7.333	1.107, 2.158	5.093	[3.791, 6.393]	<0.0001	<0.0001	0.751
100	45%	8.620, 8.430	2.163, 3.066	6.457	[4.873, 8.051]	<0.0001	<0.0001	0.800
100	60%	12.020, 10.111	2.620, 3.257	9.400	[7.432, 11.296]	<0.0001	<0.0001	0.951
100	75%	12.410, 12.256	3.321, 3.473	9.089	[6.781, 11.410]	<0.0001	<0.0001	0.774

Table 1
Numeric inference performance comparison between the Query-Point (QP) and Learned Regression (LR) approaches across different population sizes and supervision removal percentages. Reported metrics include MAE, standard deviation, MAE difference (Δ), confidence intervals (CI), Statistical tests (paired t-test & Wilcoxon text), and Cohen's d_z.

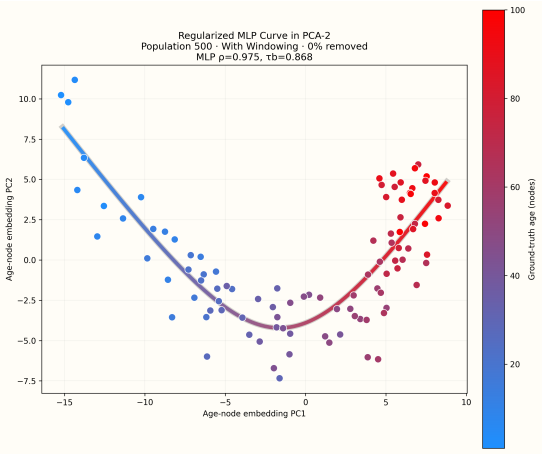
when population is at 500 and the highest was at population size of 100 at 60% removal rate. Overall, these numbers are giving a good indication that our approach was successful in preserving the structure we initially gave. In addition, the Confidence Intervals (CI) also indicate stable differences adding on to the overall success rate. That is also backed up by the statistical tests and Cohen's d_z ; showcasing robust results.

To summarize, the consistently low LR errors suggest that the latent representations preserve enough information to accurately retrieve numeric values even under significant supervision loss, while the increasing error gap as supervision decreased shows that the learned regression approach is significantly more resilient to missing age relations than direct geometric query-point inference.

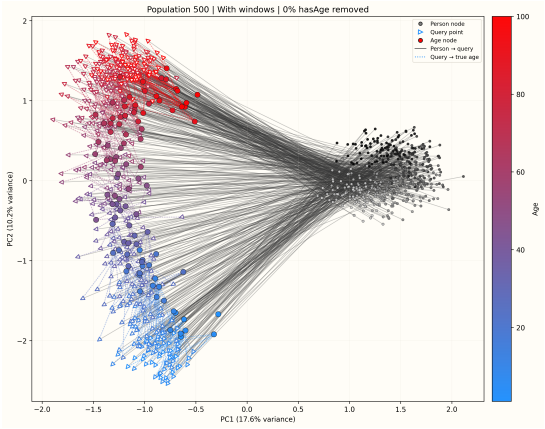
These findings support H 3, as the learned representations enabled accurate numeric inference under varying levels of supervision restriction. The consistently lower LR MAE, increasing ΔMAE advantage over QP, and statistically significant differences across removal settings indicate that the latent space preserves recoverable numeric information beyond explicitly observed *hasAge* relations.



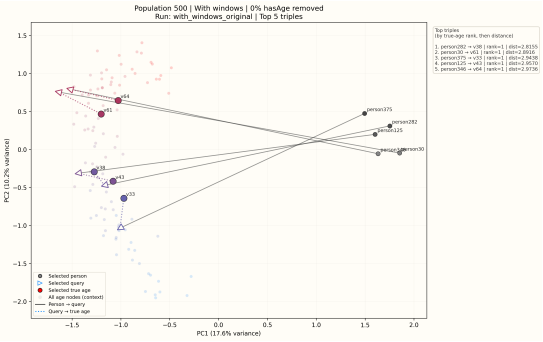
(a) PC1 of the age embeddings plotted against the original instantiated age values discarded after KG materialization.



(b) PC1 and PC2 of the age embeddings with an MLP curve, illustrating the recoverable structure in the embedding space.



(c) PCA visualization of the learned age embedding space showing the global organization of the projected representations.



(d) Top five nearest embedding neighbors for age entities, demonstrating local preservation of age similarity relationships.

Figure 4: Visualizations of the learned age embedding space.

Figure 4c and Figure 4d, give us a closer look in the mappings of the person groups and their projected age targets. The top 5 triples provide us an even closer look, where we can see that they land very close to their ground truth age. Therefore, we validate even more the geometric inference methodology we adopted.

Taken together, these results provide two complementary forms of evidence that the materialized numerical knowledge is preserved in the MuRE representations. First, the PCA projections, rank correlations, and local neighborhood structure make the monotonic organization directly observable in the embedding geometry. Second, the learned regression model recovers the corresponding numerical values with low person-level prediction error, including when direct `hasAge` supervision is progressively reduced. The encoded knowledge is therefore not only visible as an ordered geometric structure, but also quantitatively recoverable through a downstream machine-learning model.

6. Conclusion & Future Work

In conclusion, this work investigated whether numerical semantics are preserved within KGEs and whether those semantics can be recoverable after training. To answer these questions, we introduced synthetic knowledge graphs that encode monotonic numerical information and hierarchical constraints,

allowing the behavior of KGEs to be evaluated under controlled conditions through different population and supervision numbers.

Overall, the experimental findings support all three hypotheses. H1 was supported, demonstrating that materializing the transitive closure of an ordering over numeric literals reduces ordering inconsistencies between the ground-truth numeric values and their corresponding latent representations. Qualitative PCA analyses in Figures 3a, 4c and 3b together with quantitative rank-correlation results in Figure 4b showed that the original numeric ordering was consistently preserved in the learned embedding space. H2 was also supported, indicating that encoding transitive and hierarchical relations before embedding training produces stable latent representations. The resulting embedding geometry showcased consistent structural organization across multiple PCA projections, suggesting that symbolic materialization influences the learned latent space beyond the embedding model alone. Finally, H3 was supported through the numerical inference experiments showcased in Table 1. Across all population sizes and supervision levels, the Learned Regression approach consistently achieved lower prediction errors than the Query-Point method, demonstrating that numeric information encoded in the latent space remains recoverable even under substantial supervision loss.

Future work will continue investigating the behavior of KGEs using these synthetic knowledge graphs while eventually extending the proposed methodology to real-world datasets. We furthermore plan to evaluate whether the preservation of numerical semantics observed in this work generalizes to domain-specific knowledge graphs in areas such as manufacturing, polymer science and healthcare, where numeric literals (e.g., physical property values or health indicator readings) and their hierarchical relationships play an important role. Ultimately, we aim to develop a deeper understanding of how different KGE models encode symbolic and numerical information, enabling more interpretable latent representations and providing guidance on selecting and designing embedding approaches for real-world applications.

7. Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly in order to: Grammar and spelling check. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

Supplemental Materials Statement: Our public repository KGE-Neurosymbolic-AI contains:

1. Code used for creating, training and visualization.
2. Figures that are in this manuscript along with extra ones that are interesting in terms of interpretation.
3. Explanations of the overall organization and visualization aspects of the work.

Acknowledgments

Antrea Christou and Cogan Shimizu acknowledge support from the National Science Foundation (NSF) under Grant #2333532; Proto-OKN Theme 3: An Education Gateway for the Proto-OKN. CS acknowledges further support from NSF Grants #2537485 & 2514820.

References

[1] C. d'Amato, N. F. Quatraro, N. Fanizzi, Injecting background knowledge into embedding models for predictive tasks on knowledge graphs, in: R. Verborgh, K. Hose, H. Paulheim, P. Champin,

M. Maleshkova, Ó. Corcho, P. Ristoski, M. Alam (Eds.), The Semantic Web - 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings, volume 12731 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 441–457. URL: https://doi.org/10.1007/978-3-030-77385-4_26. doi:10.1007/978-3-030-77385-4_26.

[2] J. Cao, J. Fang, Z. Meng, S. Liang, Knowledge graph embedding: A survey from the perspective of representation spaces, ACM Comput. Surv. 56 (2024). URL: https://doi.org/10.1145/3643806. doi:10.1145/3643806.

[3] H. Paulheim, N. Hubert, J. Portisch, What Do Knowledge Graph Embeddings Learn to Represent?, IOP Press Ebooks, 2025, pp. 166 – 195. doi:10.3233/FAIA250206.

[4] A. Christou, Evaluating knowledge graph embeddings, in: Proceedings of the ESWC 2026 PhD Symposium, 2026. In press.

[5] B. Dave, C. Shimizu, Towards understanding the impact of schema on knowledge graph embeddings, in: S. Jain, N. Mihindukulasooriya, V. Janev, C. M. Shimizu (Eds.), Semantic Intelligence, Springer Nature Singapore, Singapore, 2024, pp. 3–10.

[6] B. Dave, A. Christou, C. Shimizu, Towards understanding the impact of graph structure on knowledge graph embeddings, in: T. R. Besold, A. d'Avila Garcez, E. Jiménez-Ruiz, R. Confalonieri, P. Madhyastha, B. Wagner (Eds.), Neural-Symbolic Learning and Reasoning - 18th International Conference, NeSy 2024, Barcelona, Spain, September 9-12, 2024, Proceedings, Part II, volume 14980 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 41–50. doi:10.1007/978-3-031-71170-1_5.

[7] A. Christou, C. Shimizu, Experiments in graph structure and knowledge graph embeddings, Neurosymbolic Artificial Intelligence (2024). In press.

[8] A. Klironomos, B. Zhou, Z. Zheng, G.-E. Mohamed, H. Paulheim, E. Kharlamov, Realite: Enrichment of relation embeddings in knowledge graphs using numeric literals, in: E. Curry, M. Acosta, M. Poveda-Villalón, M. van Erp, A. Ojo, K. Hose, C. Shimizu, P. Lisena (Eds.), The Semantic Web, Springer Nature Switzerland, Cham, 2025, pp. 41–58.

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

[10] N. Gelal, C. Snow, K. M. Jagodnik, A. Rios, H. K. McGinty, Lit-graph: Evaluating deep vs. shallow graph embeddings for high-quality text recommendation in domain-specific knowledge graphs, 2026. URL: https://arxiv.org/abs/2602.07307. arXiv:2602.07307.

[11] N. E. H. Ben Chaabene, Z. Selmi, H. Hammami, Transexplain-ld with neuro-symbolic embeddings, in: Proceedings of the 18th International Conference on Agents and Artificial Intelligence - Volume 1: EAA, INSTICC, SciTePress, 2026, pp. 946–957. doi:10.5220/0014655500004052.

[12] G. Munda, D. Gromann, T. Heinzle, Vector Space Transformations to Uncover Knowledge Graphs in Neural Language Models, IOS Pres, 2025. doi:10.3233/FAIA250204.

[13] N. Hubert, P. Monnin, A. Brun, D. Monticolo, Treat different negatives differently: Enriching loss functions with domain and range constraints for link prediction, in: The Semantic Web: 21st International Conference, ESWC 2024, Hersonissos, Crete, Greece, May 26–30, 2024, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2024, p. 22–40. URL: https://doi.org/10.1007/978-3-031-60626-7_2. doi:10.1007/978-3-031-60626-7_2.

[14] I. Balažević, C. Allen, T. Hospedales, Multi-relational poincaré graph embeddings, Curran Associates Inc., Red Hook, NY, USA, 2019.

[15] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13, Curran Associates Inc., Red Hook, NY, USA, 2013, p. 2787–2795.

[16] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in: B. Bonet, S. Koenig (Eds.), Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., AAAI Press, 2015, pp. 2181–2187.

[17] C. Spearman, The proof and measurement of association between

two things, International Journal of Epidemiology 39 (2010) 1137–1150. URL: https://doi.org/10.1093/ije/dyq191. doi:10.1093/ije/dyq191. arXiv:https://academic.oup.com/ije/article-pdf/39/5/1137/18481215/dyq191.pdf.

[18] M. G. KENDALL, A new measure of rank correlation, Biometrika 30 (1938) 81–93. URL: https://doi.org/10.1093/biomet/30.1-2.81. doi:10.1093/biomet/30.1-2.81. arXiv:https://academic.oup.com/biomet/article-pdf/30/1-2/81/423380/30-1-2-81.pdf.

[19] Student, The probable error of a mean, Biometrika 6 (1908) 1–25. doi:10.2307/2331554.