

Native Logical and Hierarchical Representations with Subspace Embeddings

Gabriel Moreira^{1,2}, Zita Marinho², Manuel Marques², João Paulo Costeira², Chenyan Xiong¹

¹Language Technologies Institute, Carnegie Mellon University

²Institute for Systems and Robotics, Instituto Superior Técnico

{gmoreira, cx}@andrew.cmu.edu, zmarinho@tecnico.ulisboa.pt, {manuel, jpc}@isr.tecnico.ulisboa.pt

Abstract

Traditional neural embeddings represent concepts as points, excelling at similarity but struggling with higher-level reasoning and asymmetric relationships. We introduce a novel paradigm: embedding concepts as linear subspaces. This framework inherently models generality via subspace dimensionality and hierarchy through subspace inclusion. It naturally supports set-theoretic operations like intersection (conjunction), linear sum (disjunction) and orthogonal complements (negations), aligning with classical formal semantics. To enable differentiable learning, we propose a smooth relaxation of orthogonal projection operators, allowing for the learning of both subspace orientation and dimension. Our method achieves state-of-the-art results in reconstruction and link prediction on WordNet. Furthermore, on natural language inference benchmarks, our subspace embeddings surpass bi-encoder baselines, offering an interpretable formulation of entailment that is both geometrically grounded and amenable to logical operations.

Code — <https://github.com/gabmoreira/subembed>

1 Introduction

Dense vector embeddings have become the bedrock of modern machine learning, underpinning systems from language models (LMs) (Devlin et al. 2019; Reimers and Gurevych 2019) and vision-language models (VLMs) (Radford et al. 2021; Li et al. 2022), to advanced retrieval augmented generation (RAG) systems (Lewis et al. 2020). By representing words, documents, images, and graph nodes as points in high-dimensional space, these representations excel at capturing nuanced similarities in a scalable manner.

Despite their widespread success, vector-based embedding paradigms inherently struggle to represent fundamental aspects of human language: compositional logic and hierarchical structure (Horn 1972). For example, representing partial orders (Vendrov et al. 2016) such as logical implications or entailment remains elusive. Similarly, capturing semantic opposition like negations (Weller, Lawrie, and Van Durme 2024; Quantmeyer, Mosteiro, and Gatt 2024; Zhang et al. 2025; Alhamoud et al. 2025) and logical conjunctions, as in “an umbrella $\wedge \neg$ it is raining” (Gokhale et al. 2020), often requires *ad-hoc* solutions lacking generalizability. This limitation is evident in recent works showing that even mod-

ern VLMs miss logical connectives, treating concepts additively, akin to bag-of-words representations (Yuksekgonul et al. 2023; Moreira et al. 2025). Such limitations impair AI systems’ ability to interpret nuanced instructions or queries in critical domains. A unified framework that can natively capture logical and hierarchical relations within a tractable embedding space remains an open and critical challenge.

To address these fundamental limitations, we propose an alternative that extends Euclidean vector representations: instead of mapping a concept to a single point, we embed it as a linear subspace in \mathbb{R}^d , the span of a set of learned vectors. This framework shifts our understanding of conceptual representation from “what a concept is” to “the set of all its possible realizations”. This enables an interpretable geometric understanding of conceptual properties:

- Generality and specificity are captured by subspace dimensionality, where higher dimensions denote broader concepts *e.g.*, animal vs. dog.
- Hierarchy is naturally modeled by subspace inclusion, where a more specific concept’s subspace is contained within a more general one.
- Logical operations are directly mapped to linear-algebraic operations: conjunction as subspace intersection, disjunction as linear sum (span), and negation as the orthogonal complement. This provides interpretable logical reasoning in the embedding space (Fig. 1).

A key advantage of our approach is its compatibility with Euclidean geometry, preserving learnability and allowing seamless integration with highly efficient dot product-based similarity search methods (Douze et al. 2025; Johnson, Douze, and Jégou 2019).

Learning such subspace representations presents a unique challenge, as the space of subspaces with varying dimensions forms a non-smooth stratified manifold. We overcome this by parameterizing subspaces with soft orthogonal projection operators, which ensure differentiability and enable end-to-end learning of subspace orientation and dimension.

We validate our framework across diverse lexical and textual entailment tasks, demonstrating its power and versatility. On WORDNET, we attain new state-of-the-art results in link prediction and reconstruction benchmarks. When transferred to HYPERLEX, a graded lexical entailment task, our embeddings significantly outperform prior methods without

requiring any task-specific tuning. For Natural Language Inference (NLI) on SNLI, our approach offers a novel geometric and interpretable formulation of textual entailment. It not only surpasses the accuracy of bi-encoder baselines but also provides direct insights into the model’s reasoning by aligning semantic entailment with geometric inclusion. In fact, by training our models for entailment (or hypernymy), we indirectly learn embeddings that are amenable to logical composition via conjunctions, disjunctions and complements. Further, as the dimensionality of the learned subspaces correlates with *generality*, our approach natively supports compression.

In summary, our key contributions are:

- A novel embedding framework that represents concepts as subspaces, enabling logical operations (conjunction, disjunction, negation).
- A spectral regularization technique that enables differentiable, end-to-end learning of subspaces with varying dimensions via projection operators, addressing a core challenge in learning structured geometric embeddings.
- A significant performance improvement on WORDNET reconstruction and hierarchy recovery and a novel interpretable geometric reasoning paradigm for Natural Language Inference, without resorting to task-specific loss functions or direct logical supervision.

2 Background

Most modern embedding methods from seminal works like Word2Vec (Mikolov et al. 2013) to advanced multimodal models such as CLIP (Radford et al. 2021) rely on a simple idea: represent entities as points (vectors) in a high-dimensional metric space $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$. The underlying premise is that similarity between entities is encoded by the inner product (or a related distance metric) between their corresponding vectors: $\mathbf{u}^\top \mathbf{v} \approx \text{sim}(u, v)$. This similarity-by-proximity paradigm often implies that a concept’s generality is captured by the magnitude of its embedding vector; the more entities a concept is similar to, the more vectors it has large inner products with, and thus, the higher its norm (Alper and Averbuch-Elor 2024). This principle extends to specialized representations, such as hyperbolic and Gaussian embeddings, where specificity correlates with distance to the origin and entropy, respectively.

Limitations of State-of-the-Art Representations. This prevalent vector-based view, while powerful for capturing co-occurrence patterns, exhibits limitations: the inner product cannot capture asymmetric relationships, such as entailment or hierarchies, without additional structural constraints or complex transformations. Recent empirical analyses have shed light on how language and vision-language encoder models represent hierarchies (Park et al. 2025; He et al. 2024) and logical constructs. Remarkably, instead of capturing formal logical structure, vector embeddings behave akin to bag-of-words representations (Yuksekgonul et al. 2023), failing to differentiate between positive and negated concepts (Gokhale et al. 2020; Singh et al. 2024; Moreira et al. 2025; Alhamoud et al. 2025). This limitation has motivated the creation of enhanced datasets and benchmarks with

explicit negations (Quantmeyer, Mosteiro, and Gatt 2024; Weller, Lawrie, and Van Durme 2024; Zhang et al. 2025).

Hyperbolic Embeddings. Hyperbolic embeddings (Nickel and Kiela 2018, 2017), as well as fully hyperbolic neural networks (Ganea, Bécigneul, and Hofmann 2018b), leverage the exponential volume growth of hyperbolic space to naturally represent hierarchical data. This geometric inductive bias enables more compact embeddings of tree-like structures and explicit modeling of transitive inclusion (Bai et al. 2021). These have been applied in diverse contexts such as hierarchical image classification (Dhall et al. 2020), structured logical multi-label prediction (Xiong et al. 2022), and safety-critical reasoning via entailment (Poppi et al. 2025). However, hyperbolic embeddings require complex Riemannian optimization routines, do not natively support logical reasoning, and their constant negative curvature makes them ill-suited for representing non-hierarchical relations (Sala et al. 2018; Moreira et al. 2024).

Partial Order Embeddings Partial order embeddings (Vendrov et al. 2016; Li, Vilnis, and McCallum 2017) model hierarchical relations by embedding entities within partially ordered spaces. Different flavors of this approach have been set forth. Positive operator embeddings (Lewis 2019) represent concepts as positive semidefinite matrices, allowing algebraic composition. Probabilistic embeddings, including Gaussian and mixture models (Vilnis and McCallum 2015; Athiwaratkun and Wilson 2018; Choudhary et al. 2021), Beta distributions (Ren and Leskovec 2020), and box lattice measures (Vilnis et al. 2018; Li et al. 2018; Ren, Hu, and Leskovec 2020), leverage convex geometric objects to capture uncertainty and logical entailment. The same principle underlies entailment cones (Zhang et al. 2021; Pal et al. 2025; Ganea, Bécigneul, and Hofmann 2018a; Yu et al. 2024), which explicitly model transitive inclusion relationships using convex cones. These approaches often encode logical entailment as geometric inclusion or overlap between distributions. Yet, they typically rely on heuristics or approximate methods for handling negation and disjunction, limiting their expressiveness.

3 Conceptual Subspace Representations

This paper presents a paradigm shift in conceptual embeddings: rather than representing a concept as a single vector $\mathbf{x} \in \mathbb{R}^d$, we represent it as a subspace $\mathcal{S} \subseteq \mathbb{R}^d$. This subspace is spanned by a set of n learnable vectors $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, making it a direct generalization that subsumes standard vector embeddings for $n = 1$. Formally, \mathcal{S} is defined as the set of all linear combinations of \mathbf{x}_i i.e.,

$$\mathcal{S} = \{\mathbf{z} \in \mathbb{R}^d \mid \mathbf{z} = \mathbf{X}\mathbf{a}, \mathbf{a} \in \mathbb{R}^n\}. \quad (1)$$

We take this subspace, not the underlying vectors \mathbf{X} , as the representation of the concept. To illustrate the core idea, consider Fig. 1. Instead of the traditional formulation, where the concept “man on a boat” is embedded as a single direction, we map it to n vectors e.g., \mathbf{x}_1 and \mathbf{x}_2 . These vectors are learned to capture distinct facets of the concept:

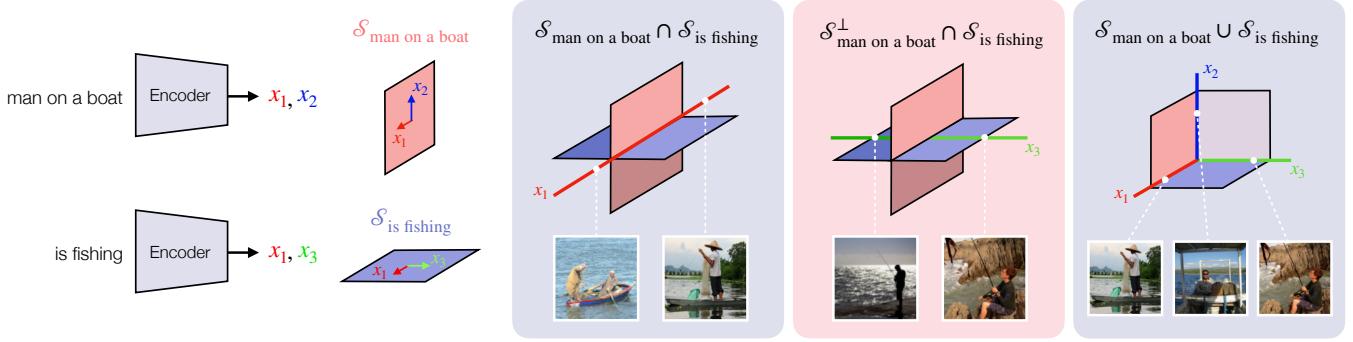


Figure 1: We embed concepts as linear subspaces of \mathbb{R}^d (left). These representations enable logical operations: subspace intersections e.g., “man on a boat” and “is fishing” (middle left); negation and composition e.g., orthogonal complement of “of man on a boat” and “is fishing” (middle right) and linear sums of subspaces, which yield a higher variance of instances (right).

x_1 might represent a “man on a boat that is fishing” while x_2 represents a “man on a boat that is not fishing”. The concept “man on a boat” is then represented by the subspace $S_{\text{man on a boat}} = \text{span}(x_1, x_2)$, encompassing all instances that align with either x_1 , x_2 , or any combination thereof, representing the manifold of all possible scenarios (Van Rijsbergen 2004; Ganter and Wille 2024), moving beyond static vector representations to geometric spaces.

3.1 Algebraic Structure of Subspaces

The power of subspaces lies in the rich algebraic structure they inherit. Any linear subspace \mathcal{S} can be represented by its orthogonal projection operator $\mathbf{P} \in \mathbb{R}^{d \times d}$, which maps any vector of \mathbb{R}^d onto \mathcal{S} . For a subspace spanned by a set of n vectors $\mathbf{X} \in \mathbb{R}^{d \times n}$, this operator is given by

$$\mathbf{P}_X := \mathbf{X}(\mathbf{X}^\top \mathbf{X})^\dagger \mathbf{X}^\top, \quad (2)$$

where \dagger is the pseudoinverse, and satisfies: $\mathbf{P}^2 = \mathbf{P}$ (idempotent) and $\mathbf{P}^\top = \mathbf{P}$ (symmetric). This projection framework directly generalizes vector embeddings: a unit-norm vector $\mathbf{x} \in \mathbb{R}^d$ defines a 1-dimensional subspace with $\mathbf{P}_X = \mathbf{x}\mathbf{x}^\top$. It further enables an interpretable mapping of logical operations to linear-algebra, addressing the limitations of state-of-the-art embeddings discussed in §1.

Subspace Lattice The set of all subspaces of \mathbb{R}^d , together with the inclusion partial order defined as

$$S_j \leq S_i \iff S_j \subseteq S_i \iff \mathbf{P}_i \mathbf{P}_j = \mathbf{P}_j, \quad (3)$$

forms the projective geometry lattice over \mathbb{R}^d denoted as $(\text{PG}(\mathbb{R}^d), \leq)$. This lattice is bounded and orthocomplemented, with greatest and least elements corresponding to the identity matrix \mathbf{I}_d (the whole space) and the null projector $\mathbf{0}$ (the zero subspace), respectively. The lattice operations of meet (conjunction), join (disjunction), and complement (negation), correspond to classical subspace operations:

1. **Meet (\wedge)**: the intersection of subspaces $S_i \cap S_j$. This represents the set of instances that belong to both concepts.
2. **Join (\vee)**: the span (linear sum) of subspaces, $S_i + S_j = \text{span}(S_i \cup S_j)$. This represents the smallest subspace containing instances of either concept.

3. **Complement (\neg)**: the orthogonal complement S^\perp . This represents everything in the ambient space that is not in the concept’s subspace.

To illustrate, consider $S_{\text{man on a boat}}$, given by $\text{span}(x_1, x_2)$, and $S_{\text{is fishing}}$, given by $\text{span}(x_1, x_3)$, from Fig. 1. Their meet (intersection) yields direction x_1 . This corresponds to the more specific concept “man on a boat that is fishing”, which is contained within both “man on a boat” and “is fishing”, thus capturing entailment relations via subspace inclusion.

These subspace operations have the following algebraic counterparts. The projection onto the join satisfies the closed-form expression $\mathbf{P}_{i \vee j} = \mathbf{P}_i + \mathbf{P}_j - \mathbf{P}_{i \wedge j}$. Negation is naturally represented as the orthogonal complement $\mathbf{P}_{\neg i} = \mathbf{I} - \mathbf{P}_i$. The projection onto the intersection of subspaces, $\mathbf{P}_{i \wedge j}$, is generally more involved. The product $\mathbf{P}_i \mathbf{P}_j$ is an orthogonal projection onto $S_i \cap S_j$ if and only if \mathbf{P}_i and \mathbf{P}_j commute. In the general case where they do not commute, $\mathbf{P}_i \mathbf{P}_j$ is merely a contraction operation. In practice, we use the approximation $\mathbf{P}_{i \wedge j} \approx \frac{1}{2} \mathbf{P}_i \mathbf{P}_j + \frac{1}{2} \mathbf{P}_j \mathbf{P}_i$.

3.2 Representing Subspaces as Smooth Projectors

While subspace representations offer a rich and interpretable geometry, their optimization poses a challenge for gradient-based learning. This difficulty arises from the non-smoothness of the space of all subspaces. Specifically, the set of rank- k subspaces of \mathbb{R}^d forms a Grassmann manifold $\text{Gr}(k, d)$, where optimization can be performed. However, to represent concepts of varying specificity, our framework requires learning subspaces with adaptive ranks (dimensions). The set of all subspaces of \mathbb{R}^d , with ranks ranging from 0 to d , is a disjoint union of Grassmannians $\bigcup_{k=0}^d \text{Gr}(k, d)$. This union forms a stratified space, with non-smooth boundaries where the rank of the subspace changes, making it thus arduous to simultaneously learn geometric orientation and the dimensionality of subspaces via gradient descent.

Smooth Projection Operators To overcome the challenges associated with learning adaptive-rank subspaces we introduce a differentiable parameterization based on a smooth relaxation of the orthogonal projection operator from Eq. (2). Given a set of n vectors $\mathbf{X} \in \mathbb{R}^{d \times n}$ (which

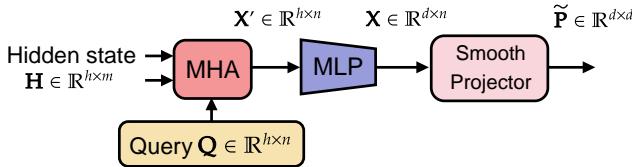


Figure 2: Subspace Projection Head (SPH) maps hidden states $\mathbf{H} \in \mathbb{R}^{h \times m}$ to soft projection operators $\tilde{\mathbf{P}} \in \mathbb{R}^{d \times d}$.

can parameterize a concept directly, or be computed from the output of an encoder model), we define a smooth projection operator via Tikhonov regularization

$$\tilde{\mathbf{P}}_X := \mathbf{X}(\mathbf{X}^\top \mathbf{X} + \Lambda)^{-1} \mathbf{X}^\top, \quad \Lambda \succ 0. \quad (4)$$

Here, Λ is a positive definite hyperparameter *e.g.*, a diagonal matrix. While $\tilde{\mathbf{P}}_X$ is not a true projector *i.e.*, not idempotent, it acts as a *soft projection* that enables differentiable learning of both orientation and effective dimensionality. This is due to its continuous spectrum: while the eigenvalues of \mathbf{P}_X from Eq. (2) are binary and thus discrete, the eigenvalues of the $\tilde{\mathbf{P}}_X$ in Eq. (4) lie within in $(0, 1)$, thus avoiding hard rank constraints and allowing dimensions to contribute gradually. As $\Lambda \rightarrow \mathbf{0}$, we recover the true projectors $\tilde{\mathbf{P}}_X \rightarrow \mathbf{P}_X$. Conversely, increasing the effect of Λ regularizes the sharpness of the projection along the principal directions of Λ . Geometrically, this changes the parameter space from a stratified manifold to a smooth manifold of positive semidefinite operators. From a Bayesian perspective, this formulation is equivalent to placing an Gaussian prior with precision Λ over the columns of \mathbf{X} .

Subspace Projection Head (SPH). To bridge our conceptual subspace representations with transformer models, we introduce the *Subspace Projection Head (SPH)*, depicted in Fig. 2. A transformer first encodes natural language inputs like ‘‘man on a boat’’ from Fig. 1 into a contextualized hidden state $\mathbf{H} \in \mathbb{R}^{h \times m}$ (where m is sequence length, h is hidden dimension). The primary function of the SPH is then to transform this hidden state \mathbf{H} into a fixed-size set of n vectors $\mathbf{X} \in \mathbb{R}^{d \times n}$ that span the concept’s subspace S , and then compute the corresponding smooth projector $\tilde{\mathbf{P}}_X$.

We distill the hidden state \mathbf{H} into a sequence-length-invariant subspace in three stages. First, we apply Multi-Head Attention (MHA) using a set of n learnable query vectors $\mathbf{Q} \in \mathbb{R}^{h \times n}$. These queries attend to \mathbf{H} (acting as keys and values), effectively pooling n embeddings $\mathbf{X}' \in \mathbb{R}^{h \times n}$,

$$\mathbf{X}' = \text{MHA}(\text{query} = \mathbf{Q}, \text{key} = \mathbf{H}, \text{value} = \mathbf{H}). \quad (5)$$

While this attention mechanism ensures invariance to sequence length m , the rank of the resulting n embeddings is still limited. We address this via a non-linear Multi-Layer Perceptron (MLP) which maps the $n h$ -dimensional vectors from the MHA output to \mathbb{R}^d . This yields the subspace matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$, whose columns define the vectors spanning the concept’s subspace

$$\mathbf{X} = \text{MLP}(\mathbf{X}'). \quad (6)$$

Finally, $\tilde{\mathbf{P}}_X$ is computed from \mathbf{X} using Eq. (4).

3.3 Training Methodology

We learn subspaces end-to-end, requiring no special initialization or training constraints beyond the inherent spectral regularization from Eq. (4). A detailed analysis of the learning dynamics is provided in the supplementary material.

Measuring Subspace Similarity and Inclusion. To quantify the degree of similarity and inclusion between subspaces using soft projections $\tilde{\mathbf{P}}$, we leverage the trace operator $\text{Tr}(\mathbf{P})$, which equals the rank of the corresponding subspace. For a soft projection, $\text{Tr}(\tilde{\mathbf{P}})$ can be interpreted as an *effective dimension*. Building on this, we define subspace similarity as the overlap between subspaces $\text{sim}(\tilde{\mathbf{P}}_i, \tilde{\mathbf{P}}_j) = \text{Tr}(\tilde{\mathbf{P}}_i \tilde{\mathbf{P}}_j)$, which can be implemented as an Euclidean dot product once vectorized *i.e.*, $\text{Tr}(\tilde{\mathbf{P}}_i \tilde{\mathbf{P}}_j) = \text{vec}(\tilde{\mathbf{P}}_i)^\top \text{vec}(\tilde{\mathbf{P}}_j)$. To quantify inclusion, we define a normalized inclusion score (Da Silva and Costeira 2009):

$$P(j | i) := \frac{\text{Tr}(\tilde{\mathbf{P}}_i \tilde{\mathbf{P}}_j)}{\text{Tr}(\tilde{\mathbf{P}}_i)} \in [0, 1]. \quad (7)$$

This score attains 1 if and only if subspace i is contained within subspace j (and if $\Lambda = \mathbf{0}$ *i.e.*, for true projectors). This formulation allows for an intuitive interpretation as a Bayes-like conditional probability: the probability of an instance belonging to subspace j given it belongs to i .

InfoNCE Loss. For similarity-based tasks, we employ a standard InfoNCE loss (van den Oord, Li, and Vinyals 2019) using the subspace overlap $\text{Tr}(\tilde{\mathbf{P}}_i \tilde{\mathbf{P}}_j)$ as similarity measure.

Link Prediction Loss. In link prediction tasks, we optimize the normalized inclusion score $P(i | j)$ directly and consider the margin loss (Vendrov et al. 2016)

$$L = \sum_{i,j \in \mathcal{P}} [\gamma_+ - P(i | j)]_+ + \sum_{i,j \in \mathcal{N}} [P(i | j) - \gamma_-]_+, \quad (8)$$

where $[\cdot]_+$ denotes the ReLU function. Here, $\gamma_+, \gamma_- \in (0, 1)$ are the positive and negative margins and \mathcal{P} and \mathcal{N} the set of positives and negatives, respectively.

NLI Classification Loss. Textual Entailment presents a unique challenge, requiring not just a measure of inclusion but also an explicit model of neutrality. For a premise p and hypothesis h , we model the relation $Y \in \{E, N, C\}$ (entailment, neutral, contradiction) as a discrete latent variable. For $Y \in \{E, C\}$, we assume the following generative process for the normalized inclusion score $S = P(h | p)$

$$S | (Y = y) \sim \text{Beta}(\alpha_y, \beta_y), \quad y \in \{E, C\}, \quad (9)$$

with $\alpha_y \leq \beta_y$ if $y = C$ and $\beta_y \leq \alpha_y$ if $y = E$. For neutrals, subspace inclusion does not provide a reliable signal. Instead, we model neutrality independently by an MLP as

$$\begin{aligned} P(Y = y | \tilde{\mathbf{P}}_p, \tilde{\mathbf{P}}_h) := \\ \sigma \left(\text{MLP} \left(\tilde{\mathbf{P}}_p, \tilde{\mathbf{P}}_h, \tilde{\mathbf{P}}_p \tilde{\mathbf{P}}_h, \tilde{\mathbf{P}}_h \tilde{\mathbf{P}}_p \right) \right), \quad y = N \end{aligned} \quad (10)$$

where $\sigma(\cdot)$ denotes the sigmoid function. Assuming uniform priors for entailment and contradiction classes, conditional on non-neutrality, we compute posterior probabilities

for $y = E$ and $y = C$, denoted $P(Y = y \mid S = s, Y \in \{E, C\})$. The final posterior probabilities for $y \in \{E, C\}$ are then derived by combining the MLP output for neutrality with the Beta posteriors for non-neutrality:

$$P(Y = y \mid \tilde{\mathbf{P}}_p, \tilde{\mathbf{P}}_h, S = s) = (1 - P(Y = N \mid \tilde{\mathbf{P}}_p, \tilde{\mathbf{P}}_h)) \cdot P(Y = y \mid S = s, Y \neq N), \quad y \in \{E, C\}. \quad (11)$$

The posterior probabilities in Eq. (10) and (11) are optimized via a cross-entropy loss.

Efficiency Considerations The key bottleneck for smooth projectors stems from the $\mathcal{O}(n^3)$ complexity of the matrix inverse in Eq. (4) and the $d \times d$ projector’s memory footprint. However, this cost is tractable for n and d up to 128, which are more than enough to achieve state-of-the-art performance (§4). Critically, for inference, once projectors are computed, similarity search leverages dot products, enabling full compatibility with highly efficient search algorithms in high-dimensional Euclidean space.

4 Experiments

In this section, we empirically validate the framework’s ability to model asymmetric relations and large-scale hierarchies, alongside its emergent capacity for logical operations. We conduct a suite of benchmarks including WORDNET (Miller 1995) reconstruction in §4.1 and link prediction in §4.2, HyperLex (Vulić et al. 2017) in §4.3, and SNLI (Bowman et al. 2015) in §4.4. All experiments were conducted on a RTX8000 GPU with 49GB of memory.

4.1 WordNet Reconstruction

In WORDNET’s reconstruction task, all edges from the full transitive closure $\text{TC}(\mathcal{G})$ of the noun and verb hypernymy hierarchies are used for training and testing. The goal is to assess the capacity of the representations to capture known hierarchical relations by providing only pairwise relations.

Training details. We parameterize each node’s subspace with a matrix $\mathbf{X}_i \in \mathbb{R}^{128 \times 128}$, initialized with entries from a zero-mean Gaussian distribution with standard deviation 0.0001. The regularizer Λ was set to $\lambda \cdot \mathbf{I}$, with $\lambda = 0.2$. For each training edge (u, v) , we sample 19 nodes $v' \neq u$ such that neither (u, v') nor (v', u) are in the train split and optimized InfoNCE using Adam (Kingma and Ba 2017), with a batch-size of 128 and learning rate of 0.0005.

Evaluation Metrics. We compute the similarity (overlap) $\text{Tr}(\tilde{\mathbf{P}}_u \tilde{\mathbf{P}}_v)$ of each edge (u, v) in the full transitive closure and rank it among the those of all node pairs that are not connected in the transitive closure $\{\text{Tr}(\tilde{\mathbf{P}}_u \tilde{\mathbf{P}}_{v'}) : (u, v') \notin \text{TC}(\mathcal{G})\}$. Given these rankings, we report the mean rank (MR) and the mean average precision (mAP). To further illustrate how our method captures the hierarchical structure of the dataset, we report the absolute value of Spearman rank-order correlation (ρ) between the ground-truth ranks of each node in the taxonomy, as defined in (Nickel and Kiela 2017), and the effective subspace dimension $\text{Tr}(\tilde{\mathbf{P}})$.

Method	Domain	mAP (\uparrow)	MR (\downarrow)	ρ (\uparrow)
Euclidean \mathbb{R}^{128}	Nouns	95.1	1.31	–
	Verbs	98.6	1.04	–
Poincaré \mathcal{P}^{10}	Nouns	86.5	4.02	58.5
	Verbs	91.2	1.35	55.1
Lorentz \mathcal{H}^{10}	Nouns	92.8	2.95	59.5
	Verbs	93.3	1.23	56.6
PG(\mathbb{R}^{128})	Nouns	98.6	1.04	68.0
	Verbs	99.9	1.00	67.0

Table 1: WORDNET **reconstruction** results. mAP = Mean Average Precision, MR = Mean Rank, ρ = Spearman correlation between taxonomy rank and subspace dimension or norm (in the case of hyperbolic embeddings).

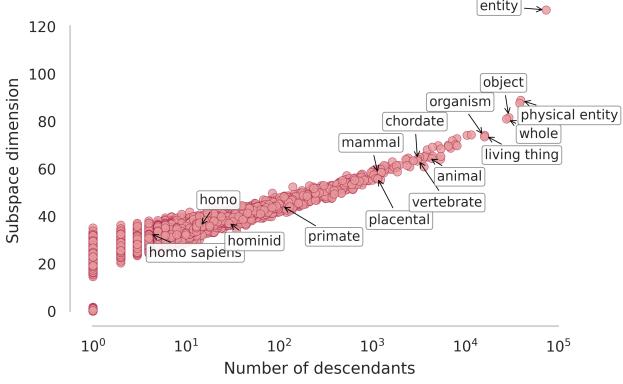
Reconstruction Results. As shown in Table 1, our subspace representations PG(\mathbb{R}^{128}) significantly outperform both Hyperbolic, \mathcal{P}^{10} and \mathcal{H}^{10} , and Euclidean embeddings \mathbb{R}^{128} . Notably, we achieve near-perfect reconstruction in the shallower verb hierarchy. The superior performance of our approach stems from its inherent geometric alignment with the underlying taxonomies. Unlike fixed-dimension point-based Euclidean or hyperbolic embeddings, which rely on pairwise distances, our model encodes the hierarchy geometrically through subspace inclusion. This allows for a much more faithful reconstruction, as evidenced by the higher rank-correlation (ρ) between the learned subspace dimensionality and the ground-truth taxonomic rank.

Further empirical evidence is provided in Fig. 3a, which illustrates how the dimension of our WORDNET noun subspaces varies with semantic generality, as measured by the number of descendants. The overlaid hypernymy chain from *homo sapiens* to the root *entity* clearly shows the subspace dimension increasing monotonically with conceptual breadth. Abstract concepts are represented by higher-rank subspaces, while specific terms like *homo sapiens* require fewer dimensions. Thus, while our model operates within \mathbb{R}^{128} , the mean effective dimension across all noun subspaces is approximately 20. This showcases the adaptability of our representations, which can dynamically allocate representational complexity based on conceptual specificity.

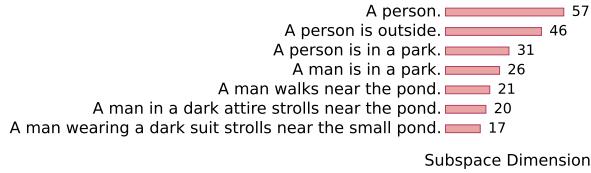
4.2 WordNet Link Prediction

In the link prediction task, we evaluate generalization from sparse supervision. We split the set of edges from the transitive closure that are not part the original graph (non-basic edges) into train (90%), validation (5%) and test (5%) using the data split from (Suzuki, Takahama, and Onoda 2019).

Training Details. To assess how the percentage of the transitive closure seen during training impacts performance, we created partial training edge coverages by randomly sampling 0%, 10%, 25% or 50% of non-basic edges, to which we append the basic edges from \mathcal{G} . We optimize the margin loss defined in (8), sampling 10 negatives for each positive.



(a) Subspace dimension vs number of descendants of our WORDNET reconstruction embeddings and dimension growth along the hypernymy path *homo sapiens* → *entity*.



(b) Subspace dimension of an entailment chain. Each sentence was encoded with our subspace model SPH $\text{PG}(\mathbb{R}^{128})$.

Figure 3: As concepts become more abstract or general, their embeddings expand into higher-dimensional subspaces.

Evaluation Metrics. We use the normalized inclusion score from Eq. (7) to evaluate edges. For each positive test edge, we consider 10 negative edges: half with a corrupted head, and half with a corrupted tail. Edges are classified as positive or negative by a threshold calibrated on the validation set. We report the classification F1-Score.

Link Prediction Results. Link prediction results are shown in Table 2. Subspace embeddings $\text{PG}(\mathbb{R}^{64})$ and $\text{PG}(\mathbb{R}^{128})$ outperform all baselines across all supervision levels. $\text{PG}(\mathbb{R}^{128})$, in particular, offers a considerable improvement when training with basic edges only (0%), owing to its larger parameter count. Notably, we achieve an F1-score of 53.4%, an improvement over the next best baseline, Order Embeddings (Vendrov et al. 2016), at 43.0%. This underscores the ability of subspace embeddings to infer hierarchical relations even from sparse supervision.

4.3 Lexical Entailment

We evaluate the capacity of our WORDNET subspace embeddings $\text{PG}(\mathbb{R}^{128})$ to capture fine-grained entailment on HYPERLEX. We use the noun subset (2,163 pairs), which provides human-annotated scores (0-10) for word pairs (u, v) , quantifying the degree to which u is a type of v . We quantify entailment using the normalized inclusion score in Eq. (7), with word sense disambiguation performed as in (Athiwaratkun and Wilson 2018), by selecting the WORDNET synset pair with maximal subspace overlap.

Method	% Non-Basic Edges			
	0%	10%	25%	50%
Euclidean \mathbb{R}^{10}	29.4	75.4	78.4	78.1
Order Embeddings \mathbb{R}^{10}	43.0	69.7	79.4	84.1
Poincaré Embeddings \mathcal{P}^{10}	29.0	71.5	82.1	85.4
Entailment Cones \mathcal{P}^{10}	32.4	84.9	90.8	93.8
Disk Embeddings \mathcal{H}^{10}	36.5	79.5	90.5	94.2
$\text{PG}(\mathbb{R}^{64})$	48.9	93.7	95.7	95.9
$\text{PG}(\mathbb{R}^{128})$	53.4	94.2	96.0	95.4

Table 2: WORDNET noun link prediction F1-Scores (\uparrow) for different percentages of non-basic edges.

	\mathbb{R}^5	\mathcal{P}^5	DOE-A	$\text{PG}(\mathbb{R}^{128})$ $\lambda=0.2$	$\text{PG}(\mathbb{R}^{128})$ $\lambda=0.6$
ρ	0.389	0.512	0.590	0.683	0.734

Table 3: Spearman’s rank correlation ρ (\uparrow) for lexical entailment on HYPERLEX using WORDNET embeddings.

As shown in Table 3, our subspace embeddings demonstrate a significantly stronger correlation with human judgments compared to existing methods. Our approach achieves Spearman’s rank correlations of 0.68 (for $\lambda = 0.2$) and 0.73 (for $\lambda = 0.6$), substantially outperforming Euclidean embeddings, Poincaré embeddings (Nickel and Kiela 2017), and Gaussian embeddings compared via KL-divergence (DOE-A) (Athiwaratkun and Wilson 2018). This notable improvement suggests that representing concepts as linear subspaces, and modeling entailment via their inclusion, offers an effective mechanism for capturing graded semantic relationships that align with human judgment.

4.4 SNLI

To demonstrate the applicability of our representations beyond word embeddings, we conducted experiments on Natural Language Inference (NLI) using the SNLI dataset. SNLI comprises 550,152 training, and 10,000 validation/test premise (p) - hypothesis (h) pairs, each annotated with one of three labels: entailment, neutral, or contradiction. For a fair comparison, we benchmarked bi-encoder baselines, using the all-miniLM-L6-v2 and mpnet-base-v2 models with a shallow MLP classifier head. We considered two common variants: $\text{MLP}(p, h)$, using concatenated premise p and hypothesis h embeddings, and $\text{MLP}(p, h, p - h)$, which also includes their difference. In our models, we map p and h to soft projection operators via our SPH module.

Training Details. We consider two regimes: 3-way, and 2-way classification (entailment vs non-entailment). We set $\Lambda = 0.05 \cdot \mathbf{I}$ in all experiments. All models were trained with a batch-size of 1024. We optimized a cross-entropy loss with label smoothing of 0.1 using Adam with a learning rate of 1e-4, decayed exponentially by a factor of 0.9. The Beta priors of our model were initialized as $(\alpha_C = 1, \beta_C = 6)$ and $(\alpha_E = 6, \beta_E = 1)$.

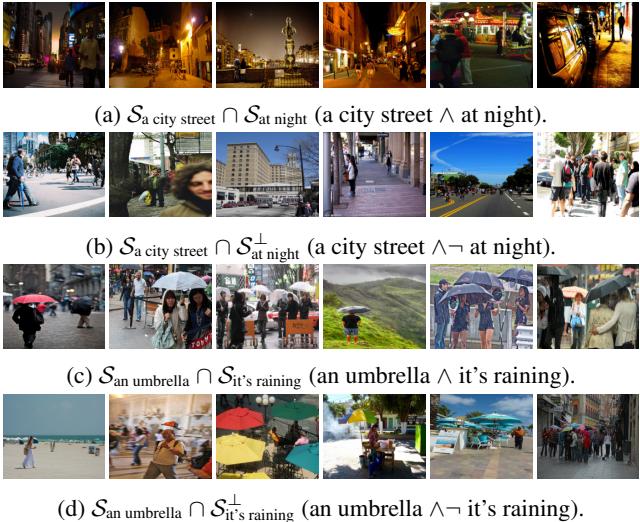


Figure 4: Flickr30k caption retrieval from logical composition of natural language queries. Queries are encoded using our mpnet-base-v2 + SPH PG(\mathbb{R}^{128}) model, finetuned on SNLI. \perp = orthogonal complement.

Results. As shown in Table 4, our approach consistently outperforms bi-encoder baselines, demonstrating that our representations’ capabilities extend beyond word embeddings. Crucially, in the 2-way setting, our method, which relies solely on a geometric inclusion for classification, consistently outperforms the opaque MLP baselines. This demonstrates that the inductive bias from the subspace structure is enough to model the complex nature of NLI, leading to superior performance with a far more interpretable mechanism. In Fig. 5 we plot the histograms of normalized inclusion scores for both 2-way and 3-way regimes, illustrating how the 3 labels correspond to interpretable degrees of inclusion, with neutrals centered around 0.5. Finally, consistent with our WORDNET findings, Fig. 3b shows that only from fine-tuning on SNLI, our model learns to map more general descriptions to larger subspaces, effectively encoding semantic generality through adaptive dimensionality.

Emergent Logical Reasoning. A key advantage of our subspace representation is its emergent logical compositionality, which arises directly from the geometry of the embeddings. Our framework, forming a subspace lattice, naturally supports operations like conjunctions, disjunctions, and negations without requiring explicit architecture or training signals. Fig. 4 provides an example illustrating this inherent compositional reasoning in a retrieval setting. For a query formed by a logical combination of concept subspaces, we retrieve images from Flickr30k (Young et al. 2014) whose caption subspaces have the largest normalized inclusion score with the composite query subspace. Each caption subspace is computed with our mpnet-base-v2 + SPH PG(\mathbb{R}^{128}) model, fine-tuned on SNLI. The results demonstrate that our subspace embeddings enable compositional retrieval, allowing for the search of novel concepts through interpretable geometric operations. Additional ex-

Method	2-way	3-way
Order Embeddings (GRU)	88.60	–
Hyperbolic Neural Network (GRU)	81.19	–
all-miniLM-L6-v2 (22.7m parameters)		
MLP(p, h)	90.48	83.63
MLP($p, h, p - h$)	91.06	84.74
SPH PG(\mathbb{R}^{32})	91.07	84.64
SPH PG(\mathbb{R}^{64})	91.02	84.62
SPH PG(\mathbb{R}^{128})	91.12	85.24
mpnet-base-v2 (109m parameters)		
MLP(p, h)	90.86	83.77
MLP($p, h, p - h$)	91.63	86.14
SPH PG(\mathbb{R}^{32})	91.74	85.54
SPH PG(\mathbb{R}^{64})	92.27	85.66
SPH PG(\mathbb{R}^{128})	92.21	86.50

Table 4: SNLI test accuracy: 2-way (entailment vs non-entailment) and 3-way classification.

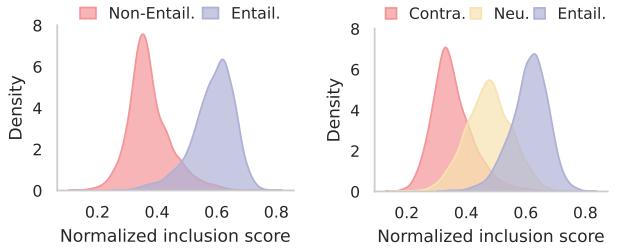


Figure 5: Normalized inclusion scores (7) for premise, hypothesis pairs from the test set of SNLI encoded with mpnet-base-v2 + SPH. Left: 2-way (entailment vs non-entailment); Right: 3-way (entailment, neutral and contradiction).

amples are provided in the supplementary material.

5 Conclusion

This paper introduced subspace embeddings, a novel paradigm for conceptual representation, addressing limitations of state-of-the-art vector representations in capturing logical structure and asymmetric relations. By representing concepts as subspaces, our framework inherently models generality via dimensionality and hierarchy through inclusion, naturally supporting compositional logic including conjunctions, disjunctions, and negations. Our comprehensive evaluation on WORDNET demonstrates unprecedented performance in reconstruction and link prediction. On SNLI, our subspace embeddings offer a unique geometric interpretation of textual entailment, outperforming bi-encoder baselines while displaying emergent propositional reasoning. Overall, this work establishes subspace embeddings as a powerful bridge between neural representations and discrete logical reasoning, opening the door for new representations that harness data’s structural nature.

References

- Alhamoud, K.; Alshammari, S.; Tian, Y.; Li, G.; Torr, P. H.; Kim, Y.; and Ghassemi, M. 2025. Vision-language models do not understand negation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 29612–29622.
- Alper, M.; and Averbuch-Elor, H. 2024. Emergent visual-semantic hierarchies in image-text representations. In *European Conference on Computer Vision*, 220–238. Springer.
- Athiwaratkun, B.; and Wilson, A. G. 2018. Hierarchical density order embeddings. In *6th International Conference on Learning Representations, ICLR 2018*.
- Bai, Y.; Ying, Z.; Ren, H.; and Leskovec, J. 2021. Modeling heterogeneous hierarchies with relation-specific hyperbolic cones. volume 34, 12316–12327.
- Bowman, S.; Angeli, G.; Potts, C.; and Manning, C. D. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 632–642.
- Choudhary, N.; Rao, N.; Katariya, S.; Subbian, K.; and Reddy, C. 2021. Probabilistic entity representation model for reasoning over knowledge graphs. volume 34, 23440–23451.
- Da Silva, N. P.; and Costeira, J. P. 2009. The normalized subspace inclusion: Robust clustering of motion subspaces. In *2009 IEEE 12th International Conference on Computer Vision*, 1444–1450. IEEE.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186.
- Dhall, A.; Makarova, A.; Ganea, O.; Pavllo, D.; Greeff, M.; and Krause, A. 2020. Hierarchical image classification using entailment cone embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 836–837.
- Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvassy, G.; Mazaré, P.-E.; Lomeli, M.; Hosseini, L.; and Jégou, H. 2025. The Faiss library. arXiv:2401.08281.
- Ganea, O.; Bécigneul, G.; and Hofmann, T. 2018a. Hyperbolic entailment cones for learning hierarchical embeddings. In *International conference on machine learning*, 1646–1655. PMLR.
- Ganea, O.; Bécigneul, G.; and Hofmann, T. 2018b. Hyperbolic neural networks. volume 31.
- Ganter, B.; and Wille, R. 2024. *Formal concept analysis: mathematical foundations*. Springer Nature.
- Gokhale, T.; Banerjee, P.; Baral, C.; and Yang, Y. 2020. Vqa-lol: Visual question answering under the lens of logic. In *European conference on computer vision*, 379–396. Springer.
- He, Y.; Yuan, M.; Chen, J.; and Horrocks, I. 2024. Language models as hierarchy encoders. volume 37, 14690–14711.
- Horn, L. R. 1972. *On the semantic properties of logical operators in English*. University of California, Los Angeles.
- Johnson, J.; Douze, M.; and Jégou, H. 2019. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3): 535–547.
- Kingma, D. P.; and Ba, J. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980.
- Lewis, M. 2019. Compositional hyponymy with positive operators. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 638–647.
- Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33: 9459–9474.
- Li, J.; Li, D.; Xiong, C.; and Hoi, S. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, 12888–12900. PMLR.
- Li, X.; Vilnis, L.; and McCallum, A. 2017. Improved Representation Learning for Predicting Commonsense Ontologies. arXiv:1708.00549.
- Li, X.; Vilnis, L.; Zhang, D.; Boratko, M.; and McCallum, A. 2018. Smoothing the geometry of probabilistic box embeddings. In *International Conference on Learning Representations*.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. volume 26.
- Miller, G. A. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11): 39–41.
- Moreira, G.; Hauptmann, A.; Marques, M.; and Costeira, J. P. 2025. Learning Visual-Semantic Subspace Representations.
- Moreira, G.; Marques, M.; Costeira, J. P.; and Hauptmann, A. 2024. Hyperbolic vs Euclidean embeddings in few-shot learning: Two sides of the same coin. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2082–2090.
- Nickel, M.; and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. volume 30.
- Nickel, M.; and Kiela, D. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, 3779–3788. PMLR.
- Pal, A.; van Spengler, M.; di Melendugno, G. M. D.; Flaboreo, A.; Galasso, F.; and Mettes, P. 2025. Compositional Entailment Learning for Hyperbolic Vision-Language Models. In *The Thirteenth International Conference on Learning Representations*.
- Park, K.; Choe, Y. J.; Jiang, Y.; and Veitch, V. 2025. The Geometry of Categorical and Hierarchical Concepts in Large Language Models. arXiv:2406.01506.
- Poppi, T.; Kasarla, T.; Mettes, P.; Baraldi, L.; and Cucchiara, R. 2025. Hyperbolic Safety-Aware Vision-Language Models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 4222–4232.

- Quantmeyer, V.; Mosteiro, P.; and Gatt, A. 2024. How and where does CLIP process negation? In *Proceedings of the 3rd Workshop on Advances in Language and Vision Research (ALVR)*, 59–72.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.
- Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992.
- Ren, H.; Hu, W.; and Leskovec, J. 2020. Query2box: Reasoning Over Knowledge Graphs In Vector Space Using Box Embeddings. In *International Conference on Learning Representations (ICLR)*.
- Ren, H.; and Leskovec, J. 2020. Beta embeddings for multi-hop logical reasoning in knowledge graphs. volume 33, 19716–19726.
- Sala, F.; De Sa, C.; Gu, A.; and Ré, C. 2018. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, 4460–4469. PMLR.
- Singh, J.; Shrivastava, I.; Vatsa, M.; Singh, R.; and Bharati, A. 2024. Learn "No" to Say "Yes" Better: Improving Vision-Language Models via Negations. arXiv:2403.20312.
- Suzuki, R.; Takahama, R.; and Onoda, S. 2019. Hyperbolic disk embeddings for directed acyclic graphs. In *International Conference on Machine Learning*, 6066–6075. PMLR.
- van den Oord, A.; Li, Y.; and Vinyals, O. 2019. Representation Learning with Contrastive Predictive Coding. arXiv:1807.03748.
- Van Rijsbergen, C. J. 2004. *The geometry of information retrieval*. Cambridge University Press.
- Vendrov, I.; Kiros, R.; Fidler, S.; and Urtasun, R. 2016. Order-Embeddings of Images and Language. arXiv:1511.06361.
- Vilnis, L.; Li, X.; Murty, S.; and McCallum, A. 2018. Probabilistic Embedding of Knowledge Graphs with Box Lattice Measures. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 263–272.
- Vilnis, L.; and McCallum, A. 2015. Word Representations via Gaussian Embedding. arXiv:1412.6623.
- Vulić, I.; Gerz, D.; Kiela, D.; Hill, F.; and Korhonen, A. 2017. HyperLex: A Large-Scale Evaluation of Graded Lexical Entailment. *Computational Linguistics*, 43(4): 781–835.
- Weller, O.; Lawrie, D.; and Van Durme, B. 2024. NevIR: Negation in Neural Information Retrieval. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2274–2287.
- Xiong, B.; Cochez, M.; Nayyeri, M.; and Staab, S. 2022. Hyperbolic embedding inference for structured multi-label prediction. volume 35, 33016–33028.
- Young, P.; Lai, A.; Hodosh, M.; and Hockenmaier, J. 2014. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2: 67–78.
- Yu, T.; Liu, T. J.; Tseng, A.; and De Sa, C. 2024. Shadow Cones: A Generalized Framework for Partial Order Embeddings. In *The Twelfth International Conference on Learning Representations*.
- Yuksekgonul, M.; Bianchi, F.; Kalluri, P.; Jurafsky, D.; and Zou, J. 2023. When and why vision-language models behave like bags-of-words, and what to do about it? arXiv:2210.01936.
- Zhang, Y.; Su, Y.; Liu, Y.; and Yeung-Levy, S. 2025. NegVQA: Can Vision Language Models Understand Negation? arXiv:2505.22946.
- Zhang, Z.; Wang, J.; Chen, J.; Ji, S.; and Wu, F. 2021. Cone: Cone embeddings for multi-hop reasoning over knowledge graphs. volume 34, 19172–19183.