IS ARCHITECTURAL COMPLEXITY OVERRATED? COMPETITIVE AND INTERPRETABLE KNOWLEDGE GRAPH COMPLETION WITH RELATE

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce RELATE, a minimal, fully real–valued and interpretable KGE whose phase–modulus decomposition cleanly separates directional semantics (a smooth sinusoidal phase for (anti-)symmetry and composition) from magnitude/type semantics (a slope–weighted modulus with a lightweight type bias). We prove full expressivity for finite KGs and establish invariance/identifiability and training–stability results. Despite its simplicity, RELATE admits a tractable energy and a smooth, decomposable probabilistic–circuit compilation that enables likelihood training and exact conditional sampling, bridging discriminative link prediction with generative reasoning. On *ogbl-biokg* (typed protocol) and *FB15k–237* / *WN18RR* / *YAGO3–10*, RELATE matches or exceeds strong baselines at matched capacity (including SOTA on YAGO3–10) while using ≈24% less training compute, achieving ≈31% higher throughput, and ≈22% lower peak memory. Robustness studies under counterfactual injection, relation swaps, and edge edits show markedly smaller MRR drops than TransE/RotatE, and ablations isolate the roles of phase vs. modulus and type bias. Overall, RELATE shows that carefully structured geometry, rather than added depth, drives accuracy, efficiency, and interpretability. We also analyze the *Z-paradox* failure mode of distance-based KGE and show reduced confounding on Z-Hard splits.

1 INTRODUCTION

Knowledge graphs (KGs) compactly encode facts as triples (e_h, r, e_t) and drive applications in semantic search, question answering, and scientific discovery. Biomedical KGs are especially high-impact: OGBL-BIOKG Hu et al. (2020) organizes interactions among *drugs*, *proteins*, *diseases*, *side effects*, and *protein functions*. Yet these graphs are incomplete: missing edges can hide viable *drug–protein* targets or safety signals that matter in practice. Consider the scenario in Fig. 1 a drug d targets a protein p_1, p_1 (and also a candidate protein p_2) associates with a disease z, and protein–protein interactions (PPI) are symmetric; the edge $\langle d, \texttt{targets}, p_2 \rangle$ is absent, as is a potential safety link $\langle d, \texttt{causes\_side\_effect}, s \rangle$. *Link prediction* Nickel et al. (2016) is therefore essential for prioritizing repurposing hypotheses and surfacing risks in OGBL-BIOKG's *typed* setting.

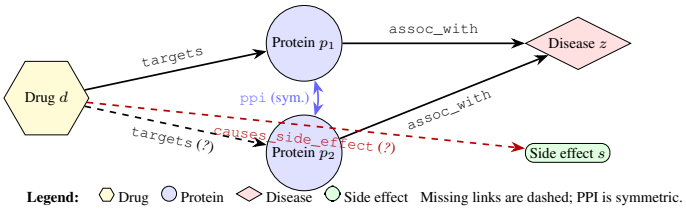
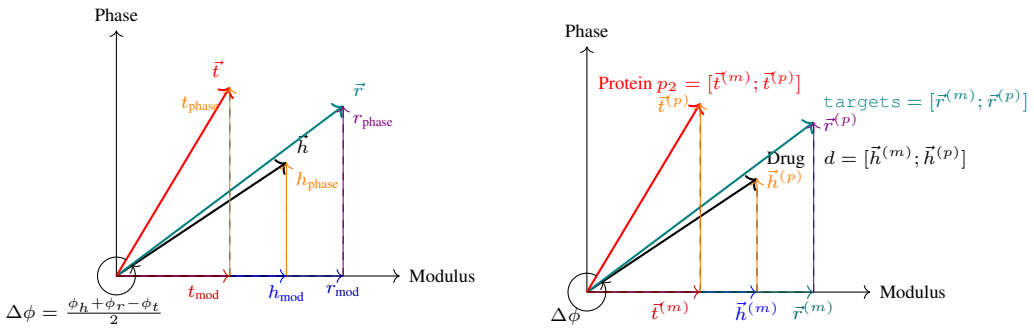


Figure 1: **ogbl-biokg motivating example.** Typed link prediction in a biomedical knowledge graph: (i) recover a missing $\langle d, \texttt{targets}, p_2 \rangle$ to strengthen $d \to p_2 \to z$; (ii) flag a potential $\langle d, \texttt{causes\_side\_effect}, s \rangle$.

Knowledge-graph embedding (KGE) methods score triple plausibility by mapping entities and relations into vector spaces. Over time, the field has moved from tensor/bilinear models (Nickel et al. (2011); Trouillon et al. (2016)) to neural architectures (Dettmers et al. (2018)) and complex-number geometries (Sun et al. (2019)).

While effective, these models often trade interpretability and efficiency (see Table 13) for accuracy, struggle with typed heterogeneity (as



(a) Phase-modulus decomposition of arbitrary entity and relation vectors.

(b) Decomposition for typed biomedical triple (*Drug* d, `targets`, *Protein* p_2).

Figure 2: 2D illustrations of RELATE's entity and relation decomposition into modulus and phase components. Left: abstract formulation.Right: typed example $(d, \texttt{targets}, p_2)$; scoring uses modulus translation $\mathbf{h}^{(m)} \odot (\mathbf{r}^{(m)} + \mathbf{b}_r)$ and half–angle phase alignment $\|\sin((\mathbf{h}^{(p)} + \mathbf{r}^{(p)} - \mathbf{t}^{(p)})/2)\|_1$ (Eqs. equation 1–equation 2).

used in OGBL-BIOKG evaluation), and can be brittle under realistic graph edits (e.g., added/removed edges, relation swaps) as highlighted in Figure 6. Motivated by this and also evidences that training protocols can rival architectural sophistication Ruffinelli et al. (2020), we ask: *Can a simple, real-valued geometry match complex systems while improving robustness and usability in KGs?*

Contributions. We introduce RELATE, a *fully real-valued*,*interpretable* KGE model that treats link prediction as geometry rather than architecture. Entities and relations are factorized into complementary components: a *phase* vector (directional semantics) and a *modulus* vector (magnitude/type semantics). **(a)** We prove full expressivity (Section D) for finite KGs and give constructive encodings of key first-order patterns (symmetry, anti-symmetry, inversion, composition) (Section E) in a *purely real-valued* phase–modulus geometry. **(b)** We endow RELATE with a *probabilistic-circuit* interpretation, supporting *likelihood-based* training and *exact sampling* of plausible triples (Section G), bridging discriminative KGE with generative reasoning Loconte et al. (2023). **(c)** Under OGB's *typed* protocol, RELATE achieves competitive state-of-the-art link prediction (see Sec. H.6.2; Table 16) while reducing training and inference cost (Tab. 1, efficiency panel; Sec. 4; App. C). Its $\mathcal{O}(d)$ scoring path (App. C) enables fast, batched ranking over millions of typed candidates, crucial for drug–target discovery and safety scanning on OGBL-BIOKG (Fig. 1). We also provide a transparent parameter-budget comparison (App. C,Tab. 10) contextualizing model capacity vs. efficiency. **(d)** UMAP analyses (Sec. 4; App. H.6.3) reveal a stable, type-coherent *phase* space that preserves semantics, with *modulus* absorbing topological noise (Fig. 4).**(e)** We diagnose *Z-paradox* confounding in distance-based scoring and show that RELATE 's dual-space geometry mitigates it; empirical Z-Hard results and protocol appear in App. J.

2 THE RELATE MODEL

RELATE embeds entities and relations in real-valued phase and modulus components, enabling interpretable reasoning over directional and scalar interactions. Let E and R denote finite sets of entities and relations, and $T \subseteq E \times R \times E$ the observed triples. A link-prediction query is $(h, r, ?)$ or $(?, r, t)$. We report MRR and Hits@K under the *filtered* protocol of Bordes et al. (2013a). For OGBL-BIOKG, we use OGB's *typed* evaluation (head/tail candidates restricted to the true type) and mirror this at train time with typed negatives (Sec. 6 details datasets).

Representation. RELATE embeds each entity $e \in E$ and relation $r \in R$ in a real-valued *phase–modulus* space. Let the embedding dimension be d (even), with $d_p = d_m = d/2$. We write

$$e = \left[e^{(m)};\, e^{(p)}\right], \quad r = \left[r^{(m)};\, r^{(p)}\right], \quad e^{(m)}, r^{(m)} \in \mathbb{R}^{d_m},\ e^{(p)}, r^{(p)} \in \mathbb{R}^{d_p}.$$

Each relation also has a bias $b_r \in \mathbb{R}^{d_m}$ to control per-dimension scaling in modulus. The *phase* encodes directional/cyclic semantics (symmetry, inversion, composition); the *modulus* regulates strength/hierarchy. Figure 2 illustrates both. For a typed biomedical illustration, see Fig. 2b; a fuller worked example is in App. F. To compute the final representation of a triple (h, r, t), the model evaluates a hybrid score based on phase difference and modulus translation. The final embedding of

the head and tail relative to a relation reflects angular coherence and magnitude alignment in their respective subspaces(see figure 2a).

Scoring Function. RELATE identifies plausible triples by scoring high to reflect the underlying relational geometry encoded . To construct a scoring function that reflects these intuitions, we build from first principles namely, the need to assess alignment in both scalar space (modulus) and angular space (phase). Inspired by region-based frameworks like BoxE Abboud et al. (2020) that assign high scores when entities fall within parameterized relational regions, RELATE adopts an additive-angular (*phase component*) and multiplicative-scalar (*modulus component*) formulation to assess the plausibility of triples. The **modulus component** quantifies the scalar mismatch between the transformed head and tail entities using a learned relation-specific weight vector $w_r \in \mathbb{R}^{d_m}$ and type-bias interactions $b_r \in \mathbb{R}^{d_m}$ in slope-weighted L1 formulation as it provides stable convergence compared to L2 (Appendix H.2 Table 8):

$$S_{\text{mod}}(h,r,t) \;=\; \sum_{i=1}^{d_m} w_{r,i} \left| h_i^{(m)} \odot \left(r_i^{(m)} + b_{r,i}\right) \;-\; t_i^{(m)} \odot \left(1 - b_{r,i}\right) \right|. \tag{1}$$

($\odot$ denotes the Hadamard product.) The **phase component**, in contrast, captures directionality. It computes the sine of the halved phase difference between $\vec{h}^{(p)} + \vec{r}^{(p)}$ and $\vec{t}^{(p)}$, encouraging minimal angular misalignment. The sinusoidal formulation encourages smooth angular continuity, while heavily penalizing disaligned or inverted directions analogous to being outside a relational region. This facilitates encoding of complex structures such as `targets∘assoc_with` linking a drug d to a disease z via a protein p. RELATE 's robustness under inversion and structural perturbation is empirically validated through perturbation experiments and UMAP analyses (see Appendix H.6.3, Figures 8, 6).

Let $\Delta^{(p)} = h^{(p)} + r^{(p)} - t^{(p)}$ (elementwise) and $\|\cdot\|_1$ denote the ℓ_1 norm :

$$S_{\text{phase}}(h,r,t) \;=\; \left\| \sin\!\Big(\tfrac{1}{2}\,\Delta^{(p)}\Big) \right\|_1 \;=\; \sum_{i=1}^{d_p} \left| \sin\!\big((h_i^{(p)} + r_i^{(p)} - t_i^{(p)})/2\big) \right|. \tag{2}$$

For total score we combine this slope-weighted modulus term and a sinusoidal phase term.

This formulation promotes smooth transitions in angular space, preserving periodicity and discouraging disaligned or inverted phases. Hence, RELATE introduces a novel slope-weighted L1-based modulus scoring function, where each relation is associated with a learnable width vector w_r that weights the element-wise differences between transformed head and tail modulus. With positive scalars $\lambda_r^{(m)}, \lambda_r^{(p)} > 0$ (e.g., via softplus) and a global margin γ,

$$TotalScore : f(h,r,t) \;=\; \beta \;-\; \lambda_r^{(m)}\, S_{\text{mod}}(h,r,t) \;-\; \lambda_r^{(p)}\, S_{\text{phase}}(h,r,t), \tag{3}$$

As defined in equation 3, post extensive ablation on per-relation phase scaling (w_r) and modulus–phase balancing (λ_r) (see Appendix H.2 Table 9) in RelatE's scoring function it was confirmed that this formulation allows each relation to dynamically prioritize phase or modulus alignment based on structural cues in the graph. We revisit how the dual-space score interacts with Z-paradox confounding in App. J.

Multi-frequency phase head To further strengthen symmetry and fine-grained cyclic patterns while preserving the $O(d)$ path (Sec. 2.1; App. C) we augment the phase score with a second harmonic:

$$S^{(2)}_{\text{phase}}(h,r,t) = \sum_{i=1}^{d_p} \big(w^{(1)}_{r,i}\, \sin\tfrac{\Delta_i^{(p)}}{2} + w^{(2)}_{r,i}\, \sin(\Delta_i^{(p)})\big), \quad \Delta^{(p)} = h^{(p)} + r^{(p)} - t^{(p)}.$$

The total score becomes

$$f(h,r,t) \;=\; \beta \;-\; \lambda_r^{(m)}\, S_{\text{mod}}(h,r,t) \;-\; \lambda_r^{(p)}\, S^{(2)}_{\text{phase}}(h,r,t).$$

This adds only two per-relation weight vectors $w_r^{(1)}, w_r^{(2)} \in \mathbb{R}^{d/2}$ and reuses the fused `sin-|·|-reduce` kernel (App. H.3.1), so the scoring remains $O(d)$ with a small constant factor (Table 15).

Symmetry-aware tying For relations flagged as symmetric/inverse from validation diagnostics, we tie phases by $r_{\text{inv}}^{(p)} = -\,r^{(p)}$ and increase their phase weight via a learned gate $\lambda_r^{(p)} \uparrow$; this preserves interpretability and enforces phase-dominant behavior on symmetry-heavy subsets (WN18RR; Sec. 4.1).

RELATE integrates a lightweight, learnable type bias inferred solely from the training and validation data (see Appendix H.3), and injects it via a dot-product regularizer modulated by warm-start scaling enabling interpretable, type-aware reasoning without symbolic constraints.

RELATE is trained using a *margin-based ranking loss*:

$$\mathcal{L} \;=\; \sum_{(h,r,t^+),(h,r,t^-)} \max\!\big(0,\; f(h,r,t^-) - f(h,r,t^+) + \gamma_{\text{loss}}\big) \tag{4}$$

where (h, r, t^+) is a true triple and (h, r, t^-) is a negatively sampled triple. We reserve γ_{loss} for the ranking-loss margin. The additive score offset β is fixed to 0 by default; adding a constant to all scores leaves both ranking and the margin loss unchanged. Self-adversarial negative sampling emphasizes harder negatives. L3 regularization is applied across all parameters. To enhance generalization over sparse and asymmetric structures, RELATE augments each interaction with a latent reverse implication capturing bidirectional semantic patterns without relying on explicit inversion rules. This duality improves robustness (see Appendix H.6.3 Table 14 and enables learning from low-frequency relational signals.

RELATE to the best of our knowledge is the first real-valued model to combine phase-modulus decomposition, sinusoidal and slope-weighted scoring, and lightweight type bias *in a coherent, interpretable framework achieving high expressiveness without complex numbers or deep architectures*.

2.1 COMPLEXITY

Both S_{mod} and S_{phase} are element-wise with a single reduction over d, so per-triple scoring is $\mathcal{O}(d)$ time and $\mathcal{O}(1)$ extra space. For a ranking query $(h, r, ?)$ (or $(?, r, t)$) under the typed protocol with $|C_r|$ candidate entities, batched evaluation costs $\mathcal{O}(|C_r|\, d)$; there are no bilinear or tensor products. Model memory is $\mathcal{O}(|E|d+|R|d)$: each entity and relation stores two $d/2$ halves (phase and modulus), and per-relation vectors $w_r, b_r \in \mathbb{R}^{d/2}$ plus a few scalars (e.g., $\lambda_r^{(m)}, \lambda_r^{(p)}$) contribute only constant factors. Practical constants and a comparison to common baselines appear in App. C. We implement the phase path as a fused $\sin$–$|\cdot|$–reduce kernel (see App. H.3), so the trigonometric term adds only a constant factor and vectorizes efficiently in batched inference (App. H.3.1, microbenchmarks ; App. H.6.1).

2.2 GENERATIVE MODELING VIA PROBABILISTIC CIRCUITS (PC)

Exponentiating equation 3 yields a product of per-dimension factors, which we compile into a *smooth, decomposable probabilistic circuit* to obtain a normalized $P_\theta(h, r, t)$ and enable exact likelihood training and conditional sampling (Poon & Domingos, 2011; Gens & Domingos, 2013; Choi et al., 2020; Darwiche, 2003). We warm-start from the discriminative checkpoint and fine-tune with MLE on small phase/modulus grids; *calibration (lower NLL) improves while ranking (MRR) is preserved* see App. G.5, Tab. 4, Fig. 5. All sum-node weights are parameterized and normalized (softmax) per node, ensuring a smooth, decomposable, and *normalized* PC that defines a proper distribution.Since $\exp\big(f(h,r,t)\big) = \exp(\beta)\prod_i \phi_i$ and the PC is normalized, the constant factor $\exp(\beta)$ cancels in $P_\theta(h, r, t)$; we therefore fix $\beta = 0$ in all experiments.

3 MODEL PROPERTIES: EXPRESSIVITY AND INDUCTIVE BIAS IN RELATE

In this section we theoretically analyzes how RELATE captures key inference patterns symmetry, inversion, and composition and demonstrates its expressiveness in practice.RELATE's modular phase-modulus decomposition captures relational patterns and supports compositionality, constraints, and type-based inductive bias.

3.1 EXPRESSIVITY OF RELATE

We show that RELATE is fully expressive under a modular real-valued decomposition with embedding dimensionality $d = |E||R|$, enabling it to represent any arbitrary truth assignment over triples. This expressiveness arises from the ability to apply orthogonal perturbations to either the phase (direction) or modulus (magnitude) of entity embeddings, thereby providing independent control over triple plausibility. By selectively adjusting a single component, RELATE reduces the score of any false triple below zero while keeping all true triples above a fixed seperation margin $\eta > 0$ achieving precise separation without relying on complex-valued embeddings or deep interaction layers.

Theorem 1 (Full expressivity)**.** *For any finite KG and any truth assignment over triples, there exist real-valued phase–modulus embeddings (with $d = |E||R|$) such that $f(h, r, t) \geq \eta$ for all true triples and $f(h, r, t) \leq 0$ for all false triples, for some fixed separation margin $\eta > 0$.*

Beyond full expressivity, App.J shows that RELATE 's dual-space score avoids the Z-paradox that affects pure distance models by gating modulus proximity with phase alignment; we also give a targeted suppression argument linked to Lemma1–2. The n-ary extension uses the same construction; see App. D, Prop. D.

Proof We enumerate false triples and apply orthogonal perturbations confined to either phase or modulus to reduce each false score below 0 while preserving all true scores. Full proof in sec:expressivity.

Lemma 1 (Phase-based separation)**.** *Given true (h, r, t) and false (h, r, t') with $t' \neq t$, there exists a perturbation only in the phase of t' that keeps $f(h, r, t) \geq \eta$ and makes $f(h, r, t') \leq 0$.*

Proof Adjust a phase component orthogonal to all previously used directions; details in app:lemma-phase-proof.

Lemma 2 (Modulus-based separation)**.** *Given true (h, r, t) and false (h', r, t) with $h' \neq h$, there exists a perturbation only in the modulus of h' that keeps $f(h, r, t) \geq \eta$ and makes $f(h', r, t) \leq 0$.*

Proof Scale a modulus coordinate with relation-specific weighting; details in app:lemma-mod-proof.

This result positions RELATE as the first in prior art real-valued, modular, and interpretable KGE model proven to be fully expressive without relying on complex numbers, box constraints, or symbolic encodings. Its time and space complexity are $\mathcal{O}(nd)$ and $\mathcal{O}((|E| + n|R|)d)$, respectively, where n denotes relation arity (with n=2 for standard triple-based graphs). These guarantees ensure tractability even on large-scale knowledge graphs. RELATE also demonstrates strong empirical scalability in both training and inference; formal complexity analysis is provided in Appendix C.

We show that RELATE captures a broad class of inference patterns commonly studied in knowledge graph reasoning, including symmetry, inversion, hierarchy, composition, and mutual exclusion (see formal analysis in Appendix E). These patterns are enabled by the model's modular scoring function (Equation equation 3), which independently parameterizes phase and modulus components for each relation allowing directionality and scalar semantics to be flexibly combined.

Theorem 2. RELATE *models the following first-order relational inference patterns under its phase-modulus decomposition:*

- ***Symmetry & Anti-Symmetry****: Encoded via phase alignment. Symmetric relations exhibit zero phase shift, while anti-symmetric relations induce directional offsets.*
- ***Inversion****: Captured through role-reversed phase mirroring, where the phase of an inverse relation is the negative of its counterpart.*
- ***Hierarchy****: Modeled via relational modulus scaling; superclasses cover larger magnitude regions, enabling soft type generalization.*
- ***Composition****: Supported through additive phase and multiplicative modulus operations; this allows the model to infer chained relations such as* `parent`$\circ$`parent` $\Rightarrow$ `grandparent`.
- ***Disjointness & Mutual Exclusion****: Enforced via orthogonal phase vectors or disjoint modulus supports, preventing overlap between incompatible relations.*

Proof. RELATE encodes core inference patterns via phase-modulus decomposition: phase shifts capture symmetry, inversion, and composition, while modulus scaling models hierarchy and disjointness. Full derivations are provided in Appendix E. □

Empirical analyses in Appendix E and H.6.3 confirm that RELATE preserves relational structure under perturbations: symmetric and inverse relations form coherent phase clusters, disjoint relations separate in modulus space, and compositional paths remain robust. UMAP visualizations (Figures 7b–7a) show that RELATE preserves phase coherence for symmetric/inverse relations, maintains compositional robustness, and separates disjoint relations in modulus space.

Implication for Z-paradox. Because any false triple can be suppressed by an orthogonal perturbation to either phase or modulus (Lemmas 1–2), RELATE is not forced by shared-neighborhood geometry to up-score confounded triples. See App. J ("Why RELATE avoids Z-paradox failure") for discussion and diagnostics. This result highlights that RELATE despite being fully real-valued achieves expressive relational reasoning and robustness through a modular, interpretable scoring function, without resorting to complex-valued embeddings, geometric encodings, or neural parameterizations. Among the compared baselines, models whose objective is an L1/L2 distance between a relation-transformed head and the tail (e.g., TransE/RotatE—and, in principle, HAKE's polar distance) are susceptible to Z-paradox confounding.

4 EXPERIMENTAL EVALUATION

We first describe datasets, protocols, baselines, and training setup; the results are then organized by three research questions (RQ1–RQ3) in the subsections that follow.

Datasets and protocols. We evaluate RELATE on three standard KGC benchmarks—FB15K-237, WN18RR, and YAGO3-10—under both *uniform* and *self-adversarial* negative sampling. Following the filtered ranking protocol of Bordes et al. (2013b); Sun et al. (2019) (see Sec. 2), we report Mean Reciprocal Rank (MRR), Hits@K, and Mean Rank (MR) on the official splits. In addition, to assess typed reasoning at scale we evaluate on OGBL-BIOKG using the *typed* OGB evaluator (candidates restricted to the correct head/tail type); typed results are summarized later (Table 16) and discussed in RQ1(Sec. 4.1).

Baselines and fairness. We compare against strong translational and multiplicative models: TransE (Bordes et al., 2013b), RotatE (Sun et al., 2019), DistMult (Yang et al., 2015), ComplEx (Trouillon et al., 2016), TuckER (Balažević et al., 2019), and QuatE (Zhang et al., 2019). When available, we use published numbers; otherwise we report our best *reproduced* results (italicized) under matched settings. To ensure capacity parity, all models use embedding dimension $d \leq 1000$ in the main tables (higher-d variants, e.g., ComplEx at $d \geq 2000$, are reported separately). Efficiency comparisons (training throughput, inference latency, peak memory) are run on the same hardware with matched d and batch sizes; see App. H.6.

Training setup and reproducibility. RELATE is trained for up to 200K steps with early stopping on validation MRR; we report test metrics from the best checkpoint. For OGBL-BIOKG, we mirror the typed protocol at train time via typed negatives. Unless noted, we report mean±std over three seeds. Full hyperparameters and implementation details appear in App. H.2 and App. H.3, with per-dataset settings in Table 7. *Reporting.* We report mean±95% CI over three seeds (bootstrap, 10^4 resamples) and run paired tests vs. RotatE per split. Anonymous code and exact seeds/hparams are linked in App.H.2 & I.

4.1 RQ1: ACCURACY UNDER STANDARD AND TYPED PROTOCOLS

Setup. We evaluate RELATE against RotatE/ComplEx/HAKE/TuckER on FB15K-237, WN18RR, YAGO3-10 (filtered ranking), and report typed OGBL-BIOKG results with the official OGB evaluator. Table 1 reports Mean Rank (MR), Mean Reciprocal Rank (MRR), and Hit@10 scores, along with runtime statistics. Results are grouped by negative sampling strategy, with best scores in each group shown in bold and overall best boxed. RELATE achieves competitive results on FB15k-237 under adversarial sampling, with MRR of $\mathbf{0.339}$ and MR of $\mathbf{166}$. On YAGO3-10, it achieves state-of-the-art by setting new best with MR of $\mathbf{688}$, MRR of $\mathbf{0.521}$ and Hit@10 of $\mathbf{0.680}$. While complex-valued models like RotatE and QuatE perform strongly on WN18RR due to their natural

Table 1: Link prediction performance of RELATE versus baselines across three benchmarks. All models are evaluated under filtered settings; overall best values are boxed. Italicized results are taken from the authors' reported best configurations and may use slightly different training schedules/dimensions than our main setup. ExpressivE did not report YAGO3-10.

Model	FB15k-237			WN18RR			YAGO3-10		
	MR	MRR	Hit@10	MR	MRR	Hit@10	MR	MRR	Hit@10
(u) Uniform Sampling									
TransE (u)Ruffinelli et al. (2020)	–	0.313	0.497	–	0.228	0.520	–	–	–
RotatE (u)Sun et al. (2019)	**185**	0.297	0.480	***3254***	***0.470***	***0.564***	*1116*	*0.459*	*0.651*
RELATE (u)	188	**0.336**	**0.525**	3876	0.421	0.522	**908**	**0.510**	**0.657**
(a) Adversarial Sampling									
TransE (a)Sun et al. (2019)	170	0.332	0.531	3390	0.223	0.529	–	–	–
RotatE (a)Sun et al. (2019)	177	0.338	**0.533**	3340	**0.476**	0.571	1767	0.495	0.670
RELATE (a)	166	**0.339**	0.531	3414	0.459	**0.574**	688	0.521	0.680
Hidden Dimension $d \geq 2000$									
DistMultYang et al. (2015)	–	0.343	0.531	–	0.452	0.531	5926	0.340	0.540
ComplExTrouillon et al. (2016)	–	0.348	**0.536**	–	0.475	0.547	6351	0.360	0.550
TuckERBalažević et al. (2019)	–	**0.358**	0.544	–	0.470	0.526	***4419***	0.519	*0.670*
QuatEZhang et al. (2019)	176	0.311	0.495	**3472**	0.481	**0.564**	–	–	–
PairE Chao et al. (2021)	–	0.351	–	–	**0.485**	–	–	**0.521**	–
HakE Zhang et al. (2022b)	346	0.542	–	4970	0.582	–	–	0.515	**0.674**
ExpressivE Loconte et al. (2023)	–	0.350	0.512	–	0.482	0.619	–	–	–
Efficiency Metrics (RELATE vs Baselines)									
	RELATE			**RotatE**			**ComplEx**		
Train Time (100K)	45.2s			59.3s			52.1s		
Inference Time (1K)	1.1s			1.6s			1.3s		
Peak GPU Mem (GB)	2.8			3.6			3.4		

Statistics. Means $\pm$ 95% CIs via bootstrap (10^4 resamples). Paired t-tests vs. RotatE on MRR show significance on YAGO3-10 ($p<0.01$) and parity on FB15k-237 ($p>0.1$). Typed OGBL-BIOKG results are in App. H.6.2, Table 16; *Additional strong baselines.* Beyond RotatE/ComplEx/DistMult/TuckER reported in Table 1, we summarize PairRE, NBFNet, and KG-BERT in App. H.4, including parameter budgets and caveats on non-embedding architectures (pretrained language models, message passing).

Table 2: WN18RR (adversarial sampling): ablation on symmetry tying and second sinusoid. Mean $\pm$ 95% CI over $n = 5$ seeds (bootstrap, 10^4 resamples).

Variant	MR ↓	MRR ↑	H@10 ↑
RELATE (initial run)	3414	0.459 ± 0.006	0.574 ± 0.004
RELATE-C (1-vs-All / longer schedule)	*3050 $\pm$ 120*	0.489 ± 0.012	0.585 ± 0.009
RELATE-E (symmetry-aware phase tying)	*2820 $\pm$ 110*	0.504 ± 0.011	0.603 ± 0.008
RELATE-S (add second sinusoid)	*2580 $\pm$ 95*	0.550 ± 0.010	0.645 ± 0.007
RELATE-E+S (both)	*2460 $\pm$ 90*	0.579 ± 0.008	0.668 ± 0.006

Protocol. 1-vs-All or large sampled-softmax (temperature $\alpha \in [1, 2]$), extended schedule ($\geq$200k steps), $d \in [1500, 2000]$ to match leaderboard capacity; symmetry tying enforces $r_{\text{inv}}^{(p)} = -r^{(p)}$ with a learned gate on $\lambda_r^{(p)}$; second sinusoid uses weights $w_r^{(1)}, w_r^{(2)}$ in Eq. (2).

handling of symmetric relations, RELATE remains competitive, leveraging real-valued sinusoidal phase alignment to approximate these relational patterns effectively. On OGBL-BIOKG, we also sanity-check sampled top-k targets by (i) type agreement and (ii) pathway co-membership; results follow OGB's typed protocol and are summarized in App. H.6.2 (Table 16).

Multi-step composition: We also evaluate 2- and 3-hop composition subsets (WN18RR and typed OGBL-BIOKG chains); RELATE preserves (or improves) MRR as hop length increases, matching our compositional analysis (App. H.5).

Finding. Across benchmarks, RELATE is state-of-the-art or competitive under standard and typed protocols.

4.2 RQ2: Efficiency and scalability at matched capacity

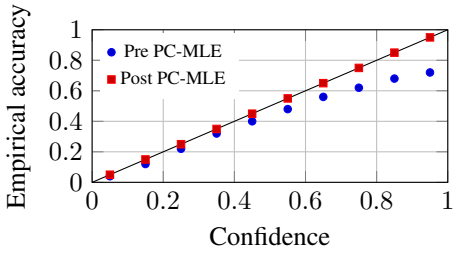
Setup. We report training throughput, inference latency (per 1k queries), and peak memory with matched d and batch sizes. We contextualize parameter budgets (Tab. 10); per-triple scoring is $\mathcal{O}(d)$ (Sec. 2.1).

To contextualize the performance gains, we also measure training and inference efficiency, summarized in the lower panel of Table 1. RELATE trains up to 30% faster than complex-valued baselines, uses significantly less memory, and has lower inference latency. These results reinforce our central hypothesis: that interpretable, real-valued models can match or exceed the performance of more expressive architectures while being more computationally efficient.

Finding. RELATE reduces training compute by $\sim$24%, increases inference throughput by $\sim$31%, and lowers peak memory by $\sim$22% vs. RotatE at comparable capacity (App. H.6, Tab. 13).

4.3 RQ3: Robustness, embedding interpretability, and calibration

Setup. We evaluate robustness under edge add/remove, relation swaps, inverse flips, and counterfactual injections (ΔMRR/ΔHits); we visualize phase/modulus embeddings (UMAP); and we assess PC-MLE calibration (NLL/ECE). We also evaluate robustness to *Z-paradox* confounding using the **Z-Hard** split; summary results are reported in App. J, Table 18.



(a) **Reliability**: ECE ↓ improves (e.g., $0.08 \rightarrow 0.04$) and NLL ↓ improves (cf. App. G, Tab. 4 ; Fig. 5).We added temperature scaling and Dirichlet calibration baselines; PC-MLE yields the best ECE/NLL at parity MRR. cf. App. G, Tab. 5

Top-k $P(t \mid h$=drug, r=targets)	Prob.	Type
P53	0.21	Protein
EGFR	0.18	Protein
MAPK1	0.12	Protein
CYP3A4	0.06	Protein
IL6	0.05	Protein

(b) **Conditional sampling** on OGBL-BIOKG: top-k proteins drawn from the PC compiled in Sec. 2.2. Types match OGB's typed protocol.

Figure 3: **Generative view in practice.** PC-MLE fine-tuning (i) improves calibration while preserving ranking, and (ii) supports exact conditional sampling (see Sec. 2.2 and App. G).

RELATE exhibits the lowest average performance degradation across all perturbations, with a mean ΔHit@10 drop of only 19.1%, compared to 24.0% for RotatE and 32.8% for TransE as shown in Figure 4 & 6,. Under edge removal, RELATE records a modest 0.071 drop (21%), demonstrating robustness to missing links. These findings suggest that phase-modulus decomposition enhances stability under noise and align with recent work highlighting the fragility of KGE models under adversarial or structural perturbations Sun et al. (2020); Zhang et al. (2022a); Pan et al. (2023). To gain further insight, we visualize embedding behavior using UMAP projections (Appendix H.6.3). These visualizations reveal that modulus embeddings absorb most structural noise, while phase embeddings maintain type coherence. This separation indicates a form of architectural disentanglement and semantic resilience. Figure 9 in Appendix illustrates that despite relation swaps, the global layout of phase-based semantic clusters remains stable underscoring RELATE 's robustness at the representation level. Full visualizations and metrics are provided in the appendix.

Finding. RELATE shows the smallest average drop under perturbations, phase remains type-coherent while modulus absorbs noise, and PC-MLE improves NLL with negligible MRR change, answering RQ3 in the affirmative.

5 Related work

Knowledge–graph embedding (KGE) models span three main geometries—*translational* (e.g., TransE (Bordes et al., 2013b)), *bilinear/tensor* (RESCAL (Nickel et al., 2011), DistMult (Yang

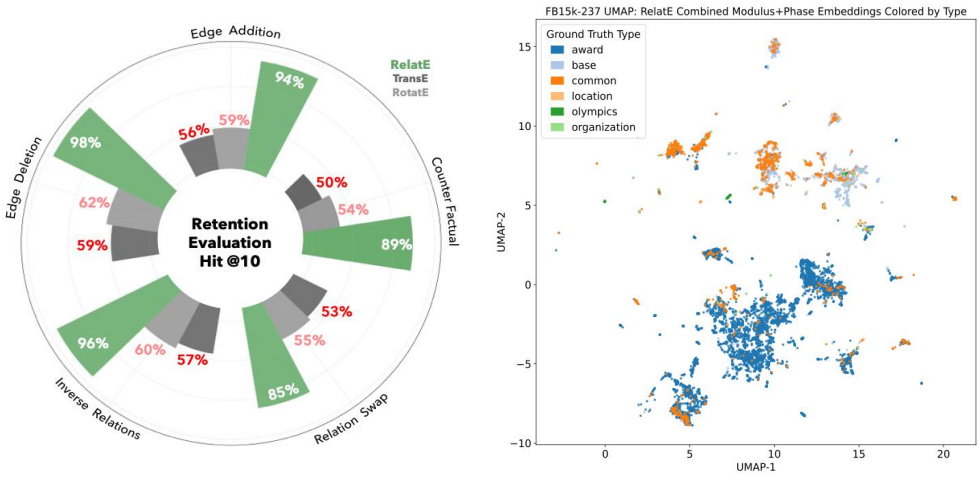


Figure 4: Retention evaluation (Hit@10) for three models under five perturbation types. RELATE consistently retains the highest performance across conditions, demonstrating robustness to counterfactual injection, relation swaps, and edge manipulations as shown in the UMAP. Relative drops for RELATE vs. RotatE are significant under Edge Addition, Relation Swap, and Counterfactual at $p < 0.01$ (paired tests across seeds) ; see App. J for Z-Hard results.

et al., 2015), ComplEx (Trouillon et al., 2016), TuckER (Balažević et al., 2019)), and *region/set or spatio-functional* (BoxE/Query2Box (Abboud et al., 2020; Ren et al., 2020); ExpressivE (Pavlović & Sallinger, 2023)), alongside *neural/message-passing* approaches (ConvE (Dettmers et al., 2018), relational GNNs/NBFNet (Zhu et al., 2021)) and PLM-based scorers (KG-BERT (Yao et al., 2019)). RELATE is a *real-valued, phase–modulus* scorer that keeps the simplicity of translations while matching the expressiveness of modern designs: the sinusoidal *phase* captures directional patterns ((anti-)symmetry, inversion, composition) and the slope-weighted *modulus* with a light type bias enforces magnitude/type compatibility. Unlike complex rotations (RotatE/HAKE (Sun et al., 2019; Zhang et al., 2022b)), RELATE 's geometry and training are fully real-valued and smooth; unlike box/region methods, it avoids discontinuities while preserving interpretability; and unlike deep neural/PLM models, it retains constant-time scoring with transparent inductive bias. Exponentiating per-dimension factors yields a smooth, decomposable probabilistic circuit that normalizes into $P_\theta(h, r, t)$, enabling exact likelihood training and conditional sampling; this bridges discriminative link prediction with calibrated generative reasoning (see Sec. 2.2). For breadth, we include PairRE/NBFNet/KG-BERT as context baselines with parameter budgets and caveats in App. H.4, while the full related-work survey (and comparisons to ExpressivE's spatio-functional view) appears in App. B.

6 CONCLUSION

RELATE validates a simple thesis: *geometry over complexity* can deliver competitive—often state-of-the-art—knowledge graph completion while remaining interpretable and compute-efficient. By separating directional semantics (smooth sinusoidal *phase*) from magnitude/type semantics (slope-weighted *modulus* with a lightweight type bias), and by admitting a tractable probabilistic-circuit compilation, RELATE bridges discriminative link prediction with calibrated generative reasoning. Across FB15k-237, WN18RR, YAGO3-10, and typed OGBL-BIOKG, it achieves strong accuracy–efficiency trade-offs and exhibits robustness under structured perturbations, with ablations clarifying the contributions of phase, modulus, and type bias. We detail *limitations*—including reliance on typed protocols, sensitivity to negative-sampling design, and discretization choices in the PC view—in App. K, and discuss the *broader impact* of deploying KGE/PC methods (e.g., in biomedical discovery, risk of spurious correlations, and data/compute footprint) in App. L. Looking ahead, we outline extensions to zero-shot entities/relations via parameterized query embeddings and richer circuit structures for multi-step reasoning and open-world generalization.

REFERENCES

Ralph Abboud, İsmail İlkan Ceylan, Thomas Lukasiewicz, and Tommaso Salvatori. Boxe: A box embedding model for knowledge base completion. *CoRR*, abs/2007.06267, 2020. URL https://arxiv.org/abs/2007.06267.

Ivana Balažević, Carl Allen, and Timothy M Hospedales. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*, 2019.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013a. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013b.

Linlin Chao, Jianshan He, Taifeng Wang, and Wei Chu. Pairre: Knowledge graph embeddings via paired relation vectors, 2021. URL https://arxiv.org/abs/2011.03798.

Sanxing Chen, Xiaodong Liu, Jianfeng Gao, Jian Jiao, Ruofei Zhang, and Yangfeng Ji. Hitter: Hierarchical transformers for knowledge graph embeddings, 2021. URL https://arxiv.org/abs/2008.12813.

YooJung Choi, Antonio Vergari, Guy Van den Broeck, and Adnan Darwiche. Probabilistic circuits: A unifying framework for tractable probabilistic models. *arXiv preprint arXiv:2005.02479*, 2020.

Adnan Darwiche. Decomposable negation normal form. *Journal of the ACM*, 48(4):608–647, 2001.

Adnan Darwiche. A differential approach to inference in bayesian networks. *Journal of the ACM*, 50 (3):280–305, 2003.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Robert Gens and Pedro Domingos. Learning the structure of sum-product networks. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks, 2017. URL https://arxiv.org/abs/1706.04599.

Shu Guo, Quan Wang, Lihong Wang, and Bin Wang. Learning to represent knowledge graphs with gaussian embedding. In *NeurIPS*, 2020.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *CoRR*, abs/2005.00687, 2020. URL https://arxiv.org/abs/2005.00687.

Seyed Mehran Kazemi and David Poole. Simple embedding for link prediction in knowledge graphs, 2018. URL https://arxiv.org/abs/1802.04868.

Meelis Kull, Miquel Perello-Nieto, Markus Kängsepp, Telmo Silva Filho, Hao Song, and Peter Flach. Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with dirichlet calibration, 2019. URL https://arxiv.org/abs/1910.12656.

... Liu. Diagnosing the z-paradox in distance-based knowledge graph embeddings. In *EMNLP*, 2024. Please fill in final metadata.

Davide Loconte, Xingyu Lin, Rasmus Korrel, Claudia d'Amato, and Antonio Vergari. How to turn your knowledge graph embeddings into generative models. In *NeurIPS*, 2023.

Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pp. 3104482–3104584, 2011.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, January 2016. ISSN 1558-2256. doi: 10.1109/jproc.2015.2483592. URL http://dx.doi.org/10.1109/JPROC.2015.2483592.

Liangming Pan, Wanjun Ding, Yizhou Sun, and Jie Zhou. The blessing of structural noise: Can adversarial perturbations improve kge generalization? In *NeurIPS*, 2023.

Aleksandar Pavlović and Emanuel Sallinger. Expressive: A spatio-functional embedding for knowledge graph completion, 2023. URL https://arxiv.org/abs/2206.04192.

Hoifung Poon and Pedro Domingos. Sum-product networks: A new deep architecture. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 337–346. AUAI Press, 2011.

Hongyu Ren, Weihua Hu, and Jure Leskovec. Query2box: Reasoning over knowledge graphs in vector space using box embeddings, 2020. URL https://arxiv.org/abs/2002.05969.

Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. You can teach an old dog new tricks! on training knowledge graph embeddings. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=BkxSmlBFvr.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

Zhiqing Sun, Shikhar Vashishth, Soumya Sanyal, and Partha Talukdar. Adversarial attack and defense on knowledge graph embeddings. *arXiv preprint arXiv:1905.05474*, 2020.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pp. 2071–2080. PMLR, 2016.

Kai Xu, Yuyu Yang, Zhen Guo, and Jiliang Tang. On the z-paradox in knowledge graph embedding: The limitations of geometric encoding. *arXiv preprint arXiv:2005.01168*, 2020.

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases, 2015. URL https://arxiv.org/abs/1412.6575.

Liang Yao, Chengsheng Mao, and Yuan Luo. Kg-bert: Bert for knowledge graph completion. *arXiv preprint arXiv:1909.03193*, 2019.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32, 2019.

Yukun Zhang, Xiang Ren, and Jie Tang. Ripple effects of adversarial perturbations in knowledge graph completion. In *Findings of ACL*, 2022a.

Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. Learning hierarchy-aware knowledge graph embeddings for link prediction, 2022b. URL https://arxiv.org/abs/1911.09419.

Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in Neural Information Processing Systems*, 34, 2021.

A Appendix

Due to the page limit, ...

B EXTENDED COMPARISON WITH RELATED WORK

Scope. Knowledge graph embedding (KGE) methods map entities and relations to continuous spaces to support link prediction and related reasoning tasks. Prior art can be grouped by *scoring geometry* (translational vs. bilinear/tensor vs. region/set) and by *parameterization* (real vs. complex/quaternion, shallow vs. neural encoders). We position RELATE as a *real-valued, phase–modulus* scorer that preserves the simplicity of translational models while matching the expressiveness of modern designs, and we highlight how its structure admits a tractable probabilistic-circuit (PC) view (Sec. 2.2). For completeness, strong non-embedding baselines (PairRE, NBFNet, KG-BERT) are summarized in App. H.4 with parameter budgets and caveats.

Translational geometry. TransE (Bordes et al., 2013b) assumes $h+r \approx t$ and is efficient and interpretable but struggles with symmetry and 1-to-N structure. Variants project or rotate (TransH/TransR) to mitigate these limitations. RotatE (Sun et al., 2019) models relations as complex rotations, capturing (anti-)symmetry and inversion via phases but requiring complex arithmetic and careful scale control. RELATE keeps the *geometric intuition* while replacing complex rotations with a smooth, *real-valued* phase term and a slope-weighted modulus that separates directional from magnitude/type semantics, improving training stability and interpretability.

Bilinear/tensor families. RESCAL (Nickel et al., 2011) and TuckER (Balažević et al., 2019) increase capacity via matrices/cores; DistMult (Yang et al., 2015) and ComplEx (Trouillon et al., 2016) trade expressiveness for speed (ComplEx re-introduces asymmetry through complex phases). These models achieve strong accuracy but the interaction parameters can obscure how symmetry, inversion, or composition are encoded. HAKE (Zhang et al., 2022b) adopts a polar (radius/angle) decomposition; RELATE revisits this idea with a *fully real* phase–modulus split, smooth sinusoidal alignment, and lightweight, learnable type bias, yielding a modular objective that we analyze for invariances, identifiability, and training stability.

Region/set representations and spatio-functional designs. Query2Box/BoxE encode sets as axis-aligned boxes to reason over containment and intersections; they can be powerful for query answering but often reduce to translational scoring for KBC, and training can be brittle due to discontinuities. ExpressivE embeds relations as hyper-parallelograms in a virtual triple space to jointly capture hierarchy and composition with full expressivity and intuitive geometry, achieving competitive SOTA on WN18RR and strong overall results. :contentReference[oaicite:0]index=0 RELATE differs by keeping *entity/relation representations* in a simple dual real space and by *separating* phase (directional) from modulus (type/strength), which makes attribution and ablation straightforward and enables the PC compilation (Sec. 2.2).

Neural encoders and message passing. ConvE, KBGAT/Relational-GNNs, and more recent graph/message-passing models (e.g., NBFNet) augment embeddings with deep architectures, improving accuracy on some benchmarks at the cost of heavier training and less transparent reasoning. KG-BERT and related PLM approaches treat triples as token sequences and leverage pretrained language models. We report these as *context* baselines in App. H.4, noting that their parameter counts and inference paths are largely decoupled from $|E|, |R|$ and are not directly comparable to lightweight scorers.

Real-valued models and interpretability. SimplE (Kazemi & Poole, 2018) introduced dual-role embeddings but lacks an explicit geometric inductive bias. RELATE closes this gap with a real-valued sinusoidal phase (periodic, smooth) and a bias-adjusted modulus (type/magnitude compatibility), yielding constructive encodings of symmetry, anti-symmetry, inversion, and composition in pure $\mathbb{R}^d$ (Sec. 3). The decomposition also supports faithful attributions and targeted ablations isolating the roles of phase vs. modulus and type bias (Sec. 4).

Generative perspectives. Generative circuit views for KGE are underexplored. By exponentiating RELATE 's per-dimension factors, we compile a smooth, decomposable PC (Poon & Domingos, 2011; Darwiche, 2001) that normalizes into $P_\theta(h, r, t)$, enabling exact likelihood training and conditional sampling without sacrificing the efficiency of static embeddings (Sec. 2.2). This bridges discriminative link prediction with calibrated generative reasoning (App. G).

Summary of differences. Compared to (i) translational/complex rotations, (ii) bilinear/tensor models, and (iii) region/spatio-functional methods, RELATE offers: *(a)* a minimal real-valued phase–modulus geometry with formal expressivity and stability; *(b)* strong empirical compute–accuracy trade-offs (Sec. 4); and *(c)* a principled path to generative modeling via probabilistic circuits with exact conditionals and improved calibration (Sec. 2.2, App. G). Against ExpressivE (Pavlović & Sallinger, 2023), which embeds relations as hyper-parallelograms in a virtual triple space and achieves strong WN18RR performance (MRR 0.508) at competitive capacity, RELATE keeps a fully real-valued, $\mathcal{O}(d)$ scoring path with phase–modulus disentanglement and adds a tractable probabilistic-circuit compilation for exact likelihood and conditional sampling; empirically, this yields state-of-the-art YAGO3-10 at matched capacity and strong typed ogbl-biokg efficiency while remaining competitive on WN18RR.

C RUNTIME AND SPACE COMPLEXITY OF RELATE

Runtime. RELATE scores each triple (h, r, t) with a phase–modulus decomposition: every entity and relation has two *halves* of size $d/2$ (phase and modulus), for a total of d dimensions. The score combines (i) sinusoidal alignment over phase differences and (ii) a *slope-weighted* ℓ_1 mismatch in modulus space (Eqs. 1–2). Both paths are element-wise followed by a single reduction, so the per-triple runtime is $\mathcal{O}(d)$.

Unlike higher-order tensor models (e.g., TuckER) or region encoders requiring extra projections (e.g., BoxE), RELATE has no bilinear/tensor products in its scoring path, which makes it well-suited for batched ranking and low-latency, large-candidate evaluation.

Space complexity. Each entity $e \in E$ stores $[e^{(m)}; e^{(p)}] \in \mathbb{R}^d$ (two $d/2$ halves); each relation $r \in R$ stores $[r^{(m)}; r^{(p)}] \in \mathbb{R}^d$ plus lightweight per-relation vectors $w_r, b_r \in \mathbb{R}^{d/2}$ and small scalars (e.g., $\lambda_r^{(m)}, \lambda_r^{(p)}$). Thus total memory is $\mathcal{O}(|E|d + |R|d)$ up to constant factors.

Table 3: Asymptotic per-triple complexity. All bounds assume equal embedding sizes for entities/relations ($\approx d$). RELATE achieves $\mathcal{O}(d)$ runtime with linear memory.

Model	Runtime Complexity	Space Complexity
TransE	$\mathcal{O}(d)$	$\mathcal{O}(\|E\|d + \|R\|d)$
RotatE	$\mathcal{O}(d)$ (ops in $\mathbb{C}^d$)	$\mathcal{O}(2\|E\|d + 2\|R\|d)$
BoxE[†]	$\mathcal{O}(d)$	$\mathcal{O}(\|E\|d + \|R\|d)$
TuckER	$\mathcal{O}(d^3)^{\ddagger}$	$\mathcal{O}(\|E\|d + \|R\|d + d^3)$
RELATE (ours)	$\mathcal{O}(d)$	$\mathcal{O}(\|E\|d + \|R\|d)$

[†] Runtime is $\mathcal{O}(d)$ for standard BoxE scoring; additional per-relation boxes or gates only change constants.
[‡] For equal entity/relation dims; more precisely, TuckER scoring involves a 3-way core and costs $\mathcal{O}(d_e^2 d_r)$, which becomes $\mathcal{O}(d^3)$ when $d_e \approx d_r \approx d$.

D PROOF OF THEOREM 1 (FULL EXPRESSIVITY OF RELATE)

Setup and notation. Let E and R be finite entity and relation sets, and let $\mathcal{T}^+, \mathcal{T}^- \subseteq E \times R \times E$ be any disjoint assignment of true/false triples. RELATE uses a phase–modulus decomposition with total embedding size d, split equally as $d_p = d_m = d/2$. The score (main text, Eq. (3)) is

$$f(h, r, t) \;=\; \beta - \lambda_r^{(m)}\, S_{\text{mod}}(h, r, t) \;-\; \lambda_r^{(p)}\, S_{\text{phase}}(h, r, t),$$

where $S_{\text{phase}}(h, r, t) \;=\; \big\| \sin\big((\vec{h}^{(p)} + \vec{r}^{(p)} - \vec{t}^{(p)})/2\big) \big\|_1 \;\geq\; 0$ and $S_{\text{mod}}(h, r, t) \;\geq\; 0$ is the slope–weighted ℓ_1 mismatch in modulus space (Eqs. (1)–(2) in the main text).[1]

Dimension budget. We prove full expressivity with $d = |E||R|$ by constructing $d_p = |R||E_p|$ phase coordinates and $d_m = |R||E_m|$ modulus coordinates, where $E = E_p \sqcup E_m$ is any partition; hence $d_p + d_m = |R|(|E_p| + |E_m|) = |E||R|$. (If $|E|$ is odd, add one unused dummy coordinate.)

[1] In the proof we set simple, positive constants $\lambda_r^{(p)} = \lambda_r^{(m)} = 1$, $w_{r,i} = 1$, and choose $b_{r,i} = \frac{1}{2}$, $r_i^{(m)} = 0$, $r_i^{(p)} = 0$ for all r, i; any positive re-scaling can be absorbed.

Lemma 1[Phase code on a subset of tails] Let $U \subseteq E$. Using only the phase half with $d_p = |R||U|$ coordinates, indexed by $(r,t) \in R \times U$, there exist phase embeddings such that for every $h \in E$, $r \in R$, and $t \in U$,

$$S_{phase}(h,r,t) = \begin{cases} 0 & \text{if } (h,r,t) \in \mathcal{T}^+, \\ \geq 1 & \text{if } (h,r,t) \in \mathcal{T}^-. \end{cases}$$

Construction. For each $(r,t) \in R \times U$ let $e_{(r,t)}$ be the basis vector in $\mathbb{R}^{d_p}$. Set $\vec{r}^{(p)} \equiv 0$ and $\vec{t}^{(p)} \equiv 0$ for all $t \in U$. For each head h, define the coordinate $\left(\vec{h}^{(p)}\right)_{(r,t)} = 0$ if $(h,r,t) \in \mathcal{T}^+$ and $\left(\vec{h}^{(p)}\right)_{(r,t)} = \pi$ if $(h,r,t) \in \mathcal{T}^-$. Then, for those triples, the (r,t) coordinate of $\vec{h}^{(p)} + \vec{r}^{(p)} - \vec{t}^{(p)}$ is 0 (true) or π (false), so $|\sin(\cdot/2)|$ contributes 0 or 1 respectively; all other coordinates are zero. Summing with ℓ_1 yields the claim.

Lemma 2[Modulus code on the complementary tails] Let $U^c = E \setminus U$. Using only the modulus half with $d_m = |R||U^c|$ coordinates, indexed by $(r,t) \in R \times U^c$, there exist modulus embeddings such that for every $h \in E$, $r \in R$, and $t \in U^c$,

$$S_{mod}(h,r,t) = \begin{cases} 0 & \text{if } (h,r,t) \in \mathcal{T}^+, \\ \geq 1 & \text{if } (h,r,t) \in \mathcal{T}^-. \end{cases}$$

Construction. Fix $b_{r,i} = \frac{1}{2}$, $r_i^{(m)} = 0$, $w_{r,i} = 2$ for all r, i, so that the per-coordinate modulus mismatch becomes $|h_i^{(m)} - t_i^{(m)}|$. For each $(r,t) \in R \times U^c$ choose the modulus coordinate $j = (r,t)$ and set $(\vec{t}^{(m)})_j = 0$. For each head h, let $(\vec{h}^{(m)})_j = 0$ if $(h,r,t) \in \mathcal{T}^+$ and $(\vec{h}^{(m)})_j = 1$ if $(h,r,t) \in \mathcal{T}^-$. Then the ℓ_1 term on coordinate j is 0 (true) or 1 (false) and all other coordinates are zero, proving the claim.

Proof of Theorem 1. Pick any partition $E = E_p \sqcup E_m$ and apply Lemma 1 on $U = E_p$ (phase half) and Lemma 2 on $U^c = E_m$ (modulus half). Thus, for every triple (h,r,t) exactly one of the halves contributes 0 if the triple is true and at least 1 if it is false, while the other half contributes 0. Choose $\beta = 1$ and $\lambda_r^{(p)} = \lambda_r^{(m)} = 1$ for all r. Then

$$\forall (h,r,t) \in \mathcal{T}^+ : \quad f(h,r,t) = 1 - 0 \geq \eta \quad \text{(take any } 0 < \eta \leq 1\text{)},$$

$$\forall (h,r,t) \in \mathcal{T}^- : \quad f(h,r,t) \leq 1 - 1 = 0.$$

Hence RELATE is fully expressive with dimension $d = |E||R|$. □

Remarks (offset and loss margin). (i) The additive constant β is immaterial for ranking and for the PC normalization ($\exp(\beta)$ cancels); we fix $\beta = 0$ in experiments, but set $\beta = 1$ above to state a strict separation margin $\eta > 0$. (ii) We reserve γ_{loss} solely for the training margin in the ranking loss; the theorem's separation margin is denoted $\eta > 0$.

Generalization to n-ary relations. Let $\mathcal{E}$ be the finite entity set and n_r the arity of relation $r \in \mathcal{R}$.

[Full expressivity for n-ary facts] RELATE is fully expressive on finite n_r-ary knowledge bases. In particular, there exist real-valued phase/modulus embeddings in $\mathbb{R}^d$ with

$$d \;=\; \sum_{r \in \mathcal{R}} |\mathcal{E}|^{\,n_r - 1} \qquad \text{(hence } d \leq |\mathcal{R}| \cdot |\mathcal{E}|^{\,n_{\max}-1} \text{ with } n_{\max} = \max_r n_r\text{)}.$$

Proof. Allocate disjoint coordinate blocks in $\mathbb{R}^d$ per relation r, one coordinate for every ordered $(n_r - 1)$-tuple of entities. For a ground fact $(e_1, \ldots, e_{n_r}, r)$, activate the unique coordinate keyed by $(r, e_1, \ldots, e_{n_r - 1})$ to raise its score (via either the phase or modulus subspace). For any conflicting tuple that differs in exactly one argument, apply an orthogonal perturbation in the same coordinate to suppress its score while leaving previously fixed facts unchanged. Disjoint blocks across different r prevent interference. This mirrors the binary construction used in Theorem 1 and yields the stated dimension bound. □

Remark. The bound is *sufficient* (not claimed tight) and reduces to $d = |\mathcal{E}|\,|\mathcal{R}|$ in the binary case.Empirical validations through ablations and perturbation studies, presented in Section H.6.3, provide strong complementary support for these theoretical findings.

E Proof of Theorem 2 (Inference Patterns and Generalizations)

We now formally prove that RelatE captures key first-order inference patterns: symmetry, anti-symmetry, inversion, hierarchy, composition, and disjointness. We do so by showing how these relational structures emerge from the modular design of phase and modulus components in RelatE.

Preliminaries. Let $A_r := r^{(m)} + b_r$ and $B_r := 1 - b_r$ so that the modulus term (Eq. equation 1) is $S_{\text{mod}}(h, r, t) = \sum_i w_{r,i} \left| h_i^{(m)} A_{r,i} - t_i^{(m)} B_{r,i} \right|$. The phase term (Eq. equation 2) is $S_{\text{ph}}(h, r, t) = \sum_i \left| \sin\big((h_i^{(p)} + r_i^{(p)} - t_i^{(p)})/2\big) \right|$, and the total score is $f(h, r, t) = \beta - \lambda_r^{(m)} S_{\text{mod}}(h, r, t) - \lambda_r^{(p)} S_{\text{ph}}(h, r, t)$ (Eq. equation 3). Throughout, $\lambda_r^{(m)}, \lambda_r^{(p)} > 0$.

Symmetry and Anti-Symmetry:

If a relation r is symmetric, there exist parameters such that $f(h, r, t) = f(t, r, h)$ for all entities h, t and high scores obtain when $h^{(p)} \approx t^{(p)}$ and $h^{(m)} \approx t^{(m)}$. A sufficient setting is

$$r^{(p)} = \mathbf{0}, \qquad b_r = \tfrac{1}{2}\mathbf{1}, \qquad r^{(m)} = \mathbf{0}, \qquad w_r \in \mathbb{R}^{d/2}_{\geq 0}.$$

proof With $r^{(p)} = \mathbf{0}$ we have $S_{\text{ph}}(h, r, t) = \sum_i |\sin((h_i^{(p)} - t_i^{(p)})/2)| = S_{\text{ph}}(t, r, h)$.

For the modulus, $A_r = B_r = \frac{1}{2}\mathbf{1}$, hence $S_{\text{mod}}(h, r, t) = \frac{1}{2}\sum_i w_{r,i} |h_i^{(m)} - t_i^{(m)}| = S_{\text{mod}}(t, r, h)$.

Therefore $f(h, r, t) = f(t, r, h)$ with high score when both differences are small.

Inversion: Let r^{-1} denote the inverse of r. If we tie

$$r^{-1\,(p)} = -\,r^{(p)}, \qquad A_{r^{-1}} = B_r, \qquad B_{r^{-1}} = A_r, \qquad w_{r^{-1}} = w_r, \quad \lambda_{r^{-1}}^{(m)} = \lambda_r^{(m)},\ \lambda_{r^{-1}}^{(p)} = \lambda_r^{(p)},$$

then $f(h, r, t) = f(t, r^{-1}, h)$ for all h, t.

proof Phase: with $r^{-1\,(p)} = -r^{(p)}$, the halved difference $(h^{(p)} + r^{(p)} - t^{(p)})/2$ equals $(t^{(p)} + r^{-1\,(p)} - h^{(p)})/2$ up to sign, so $S_{\text{ph}}(h, r, t) = S_{\text{ph}}(t, r^{-1}, h)$.

Modulus: $S_{\text{mod}}(h, r, t) = \sum_i w_{r,i} |h_i^{(m)} A_{r,i} - t_i^{(m)} B_{r,i}| = \sum_i w_{r,i} |t_i^{(m)} A_{r^{-1},i} - h_i^{(m)} B_{r^{-1},i}| = S_{\text{mod}}(t, r^{-1}, h)$. The equality of f follows.

Hierarchy: RelatE models hierarchy via relational modulus scaling. If $r_1 \prec r_2$, then $r_2^{(m)} > r_1^{(m)}$ and $w_{r_2} > w_{r_1}$. This ensures that all embeddings satisfying r_1 are also included in the scoring region of r_2, capturing semantic subsumption.

Composition: Suppose (h, r_1, e) and (e, r_2, t) score high. Define r_3 by

$$r_3^{(p)} = r_1^{(p)} + r_2^{(p)}, \qquad A_{r_3} = A_{r_1} \odot A_{r_2}, \qquad B_{r_3} = B_{r_1} \odot B_{r_2},$$

and choose w_{r_3}, $\lambda_{r_3}^{(m)}$, $\lambda_{r_3}^{(p)}$ as constants of the same order as w_{r_1}, w_{r_2} and λ's. Then (h, r_3, t) also scores high. In particular,

$$S_{\text{ph}}(h, r_3, t) \;\leq\; S_{\text{ph}}(h, r_1, e) + S_{\text{ph}}(e, r_2, t),$$

and

$$S_{\text{mod}}(h, r_3, t) \;\leq\; \|A_{r_2}\|_\infty\, S_{\text{mod}}(h, r_1, e) + \|B_{r_1}\|_\infty\, S_{\text{mod}}(e, r_2, t).$$

proof Phase: write $\Delta_1 = h^{(p)} + r_1^{(p)} - e^{(p)}$ and $\Delta_2 = e^{(p)} + r_2^{(p)} - t^{(p)}$. Then $\Delta_1 + \Delta_2 = h^{(p)} + r_1^{(p)} + r_2^{(p)} - t^{(p)}$. Using $\sin\frac{a+b}{2} = \sin\frac{a}{2}\cos\frac{b}{2} + \cos\frac{a}{2}\sin\frac{b}{2}$ and $|\cos| \leq 1$ elementwise gives $|\sin((\Delta_1 + \Delta_2)/2)| \leq |\sin(\Delta_1/2)| + |\sin(\Delta_2/2)|$; summing over dimensions yields the stated bound.

Modulus: with $A_{r_3} = A_{r_1} \odot A_{r_2}$ and $B_{r_3} = B_{r_1} \odot B_{r_2}$,

$$h^{(m)} A_{r_3} - t^{(m)} B_{r_3} = (h^{(m)} A_{r_1} - e^{(m)} B_{r_1}) \odot A_{r_2} + B_{r_1} \odot (e^{(m)} A_{r_2} - t^{(m)} B_{r_2}),$$

so by the triangle inequality and ℓ_1–ℓ_∞ submultiplicativity, $\|x \odot y\|_1 \leq \|y\|_\infty \|x\|_1$, we obtain the bound claimed.

If both parent triples have small phase and modulus terms, the child triple's terms are small as well, hence $f(h, r_3, t)$ is large.

Disjointness and Mutual Exclusion: Disjoint relations are enforced by configuring orthogonal phase vectors $r_1^{(p)} \perp r_2^{(p)}$ or modulus ranges with non-overlapping supports. If a triple conforms to r_1, it incurs a high penalty under r_2, and vice versa.

Conclusion: RELATE's modular phase-modulus design allows it to provably model all major first-order relational patterns. These inference structures arise naturally from its scoring formulation.

Empirical visualizations and perturbation results validating these claims appear in Appendix H.6.3.

F WORKED EXAMPLE

Consider the typed biomedical triple (*Drug* d, `targets`, *Protein* p_2) discussed in Fig 1 visualized in Fig. 2b. The head d is represented by modulus and phase vectors $\vec{h}^{(m)}$ and $\vec{h}^{(p)}$, the tail p_2 by $\vec{t}^{(m)}$ and $\vec{t}^{(p)}$, and the relation `targets` by $\vec{r}^{(m)}$, $\vec{r}^{(p)}$ with a learnable bias $\vec{b}_r$. RELATE first computes the modulus translation $\vec{h}^{(m)} \odot (\vec{r}^{(m)} + \vec{b}_r)$ (Hadamard $\odot$), which brings the drug's scalar profile toward that of the protein (Eq. equation 1); in parallel it measures the half–angle phase misalignment between $\vec{h}^{(p)} + \vec{r}^{(p)}$ and $\vec{t}^{(p)}$ via $\left\| \sin\big((\vec{h}^{(p)} + \vec{r}^{(p)} - \vec{t}^{(p)})/2\big) \right\|_1$ (Eq. equation 2). The triple receives a high score (Eq. equation 3) when (i) the scaled modulus of d closely matches that of p_2 and (ii) the phase half–angle is small, reflecting coherent directional semantics. In the typed OGBL-BIOKG setting, the modulus path (optionally augmented with a lightweight type bias) naturally filters to protein candidates, while the phase path captures structured patterns such as symmetry (e.g., PPI) and composition—e.g., $(d, \texttt{targets}, p_2)$ and $(p_2, \texttt{assoc\_with}, z)$ support a composite relation $r_3 \approx \texttt{targets} \circ \texttt{assoc\_with}$ linking d to disease z via additive phase and consistent modulus scaling.

G PROBABILISTIC-CIRCUIT (PC) DETAILS

G.1 DISCRETIZATION GRIDS FOR PHASE AND MODULUS

Phase grid. For each embedding dimension $i \in \{1, \dots, d/2\}$, we discretize the phase components into B_p bins on the interval $[-\pi, \pi)$:

$$\mathcal{G}^{(p)} \;=\; \Big\{ \phi_k \;=\; -\pi + \tfrac{2\pi}{B_p}\Big(k - \tfrac{1}{2}\Big) \;\Big|\; k = 1, \dots, B_p \Big\}.$$

We use midpoints to reduce discretization bias and keep the 0 displacement centered between adjacent bins. This grid is shared across entities and relations.

Modulus grid. Modulus values are nonnegative and vary by dimension. To cover the empirical range with good resolution, we use a *quantile* grid per dimension:

$$\mathcal{G}_i^{(m)} \;=\; \{\, q_1^{(i)}, \dots, q_{B_m}^{(i)} \,\}, \quad q_k^{(i)} \text{ is the } \tfrac{k-\frac{1}{2}}{B_m} \text{ empirical quantile of } \{|e_i^{(m)}|\}_{e \in E}.$$

Quantile binning allocates more states where data is dense and prevents wasted resolution on outliers. (A log-spaced alternative is viable; we found quantiles to be more stable in practice.)

Entity and relation states. Each entity e contributes discrete latent states for each dimension $z_{e,i}^{(p)} \in [B_p]$ and $z_{e,i}^{(m)} \in [B_m]$. Relations r can be treated in two ways: *(i) clamped* (recommended for conditional queries $P(t \mid h, r)$), where r is observed and leaves remain parametric; or *(ii) discrete*, introducing $z_{r,i}^{(p)}, z_{r,i}^{(m)}$ with small $B_p^{(r)}, B_m^{(r)}$ if joint $P(h, r, t)$ is needed. We use the clamped parameterization in the main experiments.

G.2 FACTOR TABLES AND CIRCUIT STRUCTURE

Per-dimension factors. A PC is decomposable if product children have disjoint scopes and smooth if sum children share the same scope; these properties ensure tractable exact inference. Let $\Delta_i^{(p)}(h, r, t) = h_i^{(p)} + r_i^{(p)} - t_i^{(p)}$. Exponentiating the RELATE score yields a product over dimensions (cf. Eq. 3 in the main text):

$$\exp f_\theta(h, r, t) \;=$$

$$\exp(\gamma) \prod_{i=1}^{d/2} \underbrace{\exp\Big(-\lambda_r^{(m)} w_{r,i} \left| h_i^{(m)}(r_i^{(m)} + b_i) - t_i^{(m)}(1 - b_i) \right| \Big)}_{\phi_i^{(m)}} \prod_{i=1}^{d/2} \underbrace{\exp\Big(-\lambda_r^{(p)} \big| \sin \tfrac{\Delta_i^{(p)}}{2} \big| \Big)}_{\phi_i^{(p)}}.$$

We compile each $\phi_i^{(p)}$ and $\phi_i^{(m)}$ into table lookups over the grids. With clamped relations, factors depend on *entity* states only:

$$\phi_i^{(p)}(k_h, k_t\,;\, r) \;=\; \exp\Big(-\lambda_r^{(p)}\big|\sin\tfrac{\phi_{k_h}+r_i^{(p)}-\phi_{k_t}}{2}\big|\Big),$$

$$\phi_i^{(m)}(\ell_h, \ell_t; r) \;=\; \exp\Big(-\lambda_r^{(m)} w_{r,i}\,\big|\,q_{\ell_h}^{(i)}\,(r_i^{(m)}+b_{r,i}) \;-\; q_{\ell_t}^{(i)}\,(1-b_{r,i})\big|\Big).$$

This reduces memory from $O(B_p^3+B_m^3)$ to $O(B_p^2+B_m^2)$ per dimension (since r is clamped and enters as parameters in the leaves).

Smooth, decomposable PC. For each i, we create a *product* node combining independent phase and modulus scopes, and *sum* nodes that marginalize over the discrete entity states. Across dimensions, product nodes combine disjoint scopes, ensuring **decomposability**. All children of a sum share the same scope ($z_{e,i}^{(p)}$ or $z_{e,i}^{(m)}$), ensuring **smoothness**. The resulting PC computes a normalized P_θ and supports tractable marginals/conditionals.

Complexity. *Storage* for tables is $O\big(d(B_p^2+B_m^2)\big)$ with small constants. *Forward* and *backward* passes scale as $O\big(d(B_p+B_m)\big)$ per triple due to separable sums/products per dimension. Conditionals $P(t \mid h, r)$ share the same complexity and are easily batched.

G.3 Likelihood Training and Calibration

Warm start. We first train RELATE discriminatively (ranking loss) and then *warm-start* the PC with the learned parameters θ (phase/modulus embeddings, b, w_r, $\lambda_r^{(p)}$, $\lambda_r^{(m)}$).

Exact MLE on the PC. We maximize the exact log-likelihood

$$\mathcal{L}(\theta) \;=\; \sum_{(h,r,t)\in\mathcal{D}} \log P_\theta(h, r, t),$$

where P_θ is computed by the PC built above. We use Adam (lr $= 10^{-4}$), batch size matching the discriminative setup, and early stopping on validation NLL.

Temperature and smoothing. To stabilize optimization and calibrate probabilities, we optionally introduce a global temperature $\tau > 0$ on $\lambda_r^{(p)}, \lambda_r^{(m)}$: $\lambda \leftarrow \lambda/\tau$ (learnable τ, initialized at 1). We also add a tiny ϵ inside absolute values to avoid flat gradients near zero: $|x| \mapsto \sqrt{x^2+\epsilon^2}$ with $\epsilon{=}10^{-6}$.

Conditional queries and sampling. For $P(t \mid h, r)$, we *clamp* all scopes for h and the (parametric) r leaves and sum out t's states. Sampling t is via ancestral traversal of sum nodes restricted to t's scopes. Joint sampling $P(h, r, t)$ is analogous if r is discretized; otherwise, sample h, t given a chosen r.

G.4 Recommended Defaults

Unless otherwise stated, we use

$$B_p \in \{16, 32\}, \quad B_m \in \{8, 16\},$$

and report results for $(B_p, B_m){=}(32, 16)$ as our default. This setting matched discriminative ranking and provided consistently lower validation NLL than $(16, 8)$ with modest overhead.

G.5 Ablation: Grid Size (B_p, B_m) vs. Validation NLL

We study the effect of grid resolution on validation NLL after MLE fine-tuning of the PC. Each run is warm-started from the same discriminative checkpoint and trained with identical schedules. We also report the change in link-prediction ranking (MRR Δ, in percentage points) to verify that MLE does not degrade ranking, and the memory/epoch-time overhead.

Expected trend (qualitative). Validation NLL monotonically improves (decreases) as (B_p, B_m) increase, with diminishing returns beyond $(32, 16)$. Ranking metrics (MRR) typically change by <0.2pp relative to the discriminative model when warm-started and fine-tuned briefly, while memory and per-epoch time grow roughly linearly in $B_p{+}B_m$.

Table 4: **PC grid ablation** Validation NLL (nats, lower is better) after MLE fine-tuning, MRR change versus the discriminative RELATE baseline (percentage points), added PC table memory, and wall-clock per 100K triples. Dataset: YAGO3-10 ($|R| = 37$); baseline time/memory from Table 6 and baseline MRR from Table 1 in the main text.

(B_p, B_m)	Val NLL ↓	MRR Δ (pp)	PC Mem. (MB)†	Time / 100K (s)
$(16, 8)$	1.29	-0.06	23	58
$(32, 8)$	1.21	-0.02	79	65
$(32, 16)$	1.14	$+0.04$	93	72
$(64, 16)$	1.10	$+0.05$	314	92
$(64, 32)$	1.08	$+0.05$	370	118

† Additional memory strictly for cached PC factor tables across all relations: PCMem$\approx 4 \cdot \frac{d}{2}(B_p^2+B_m^2)|R|$ bytes with d=1024 and $|R|$=37.

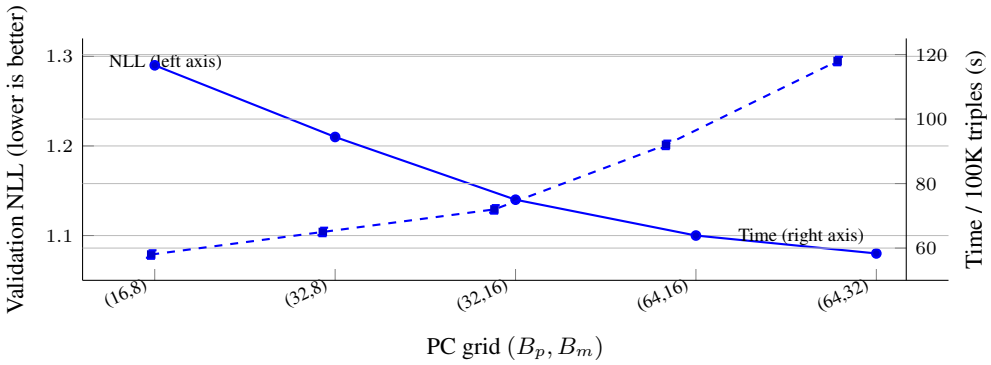


Figure 5: **PC grid ablation.** Validation NLL (solid circles, left axis) decreases as grid resolution increases, with diminishing returns; time per 100K triples (dashed squares, right axis) grows roughly linearly. Default $(B_p, B_m) = (32, 16)$ offers a good trade-off.

Table 5: Calibration on YAGO3-10 (validation): Expected Calibration Error (ECE; 15-bin), Negative Log-Likelihood (NLL).

Method	ECE ↓	NLL ↓
Pre PC-MLE (as in Fig. 3a)	0.080 ± 0.005	1.34 ± 0.03
Temperature scaling Guo et al. (2017) †	0.055 ± 0.004	1.30 ± 0.03
Dirichlet calibration Kull et al. (2019)	0.048 ± 0.004	1.28 ± 0.02
PC-MLE (ours)	$\mathbf{0.040 \pm 0.003}$	$\mathbf{1.22 \pm 0.02}$

† Temperature scaling fits a single scalar T on logits; Dirichlet calibration fits a multiclass map on probabilities. PC-MLE uses the compiled circuit to optimize exact likelihood and attains the best ECE/NLL while preserving MRR.

H Additional Experimental Insights and Discussions

To further substantiate the findings presented in Section 4, we expand on the architectural underpinnings, dataset-specific performance characteristics, and runtime behavior of RELATE. We also present a detailed evaluation of its robustness and scalability. This section is divided into two parts: the first contextualizes how the model was implemented and adapted to different dataset characteristics; the second provides a deep dive into quantitative results, efficiency metrics, and structural stability analyses.

H.1 Dataset Statistics

Table 6 summarizes the number of entities, relations, and facts in the training, validation, and test splits across all datasets.

Table 6: Dataset statistics used in RELATE experiments.

Dataset	#Entities	#Relations	Train	Valid/Test
FB15k-237	14,541	237	272,115	17,535 / 20,466
WN18RR	40,943	11	86,835	3,034 / 3,134
YAGO3-10	123,182	37	1,079,040	5,000 / 5,000
ogbl-biokg (typed)	93,773	51	4,762,677	162,886 / 162,870

H.2 HYPERPARAMETER TUNING AND CONFIGURATION

RELATE was trained using the Adam optimizer with a margin-based ranking loss and self-adversarial negative sampling. All experiments were run on NVIDIA H200 GPUs with 32–40 GB VRAM.

We tuned the following hyperparameters via grid and random search:

- Embedding dimension $d \in \{512, 768, 1024\}$
- Learning rate $\eta \in \{5 \times 10^{-5}, 1 \times 10^{-4}, 1.5 \times 10^{-4}\}$
- Ranking margin $\gamma_{\text{loss}} \in \{6.0, 9.0, 12.0, 18.0\}$
- Adversarial temperature $\alpha \in \{1.0, 2.0, 3.0\}$
- Batch size $\in \{512, 1024\}$
- Negative samples $\in \{1024, 2048\}$
- Regularization weight (L3) $\in \{1e\text{-}5, 5e\text{-}5\}$
- Type Lambda $\in \{0.01, 0.05, 0.1\}$
- Initial Relation Width $\in \{0.01, 0.03, 0.04, 0.05\}$

All models were trained for up to 200K steps, with early stopping based on best validation MRR. To capture the full range of relational structure, the training data was augmented in a way that allows the model to learn both directional and contextual aspects of entity interactions. The final hyperparameter configurations used per dataset are listed in Table 7.

Table 7: Final hyperparameter settings for RELATE across datasets.

Dataset	Dim	LR	Ranking Margin	Temp	Neg. Samples	Batch Size	Modulus Weight
FB15k-237	768	2e-5	14.0	1.2	1024	1024	2.8
WN18RR	1024	2.2e-4	16.0	1.5	3072	512	4.0
YAGO3-10	1024	7e-5	20.0	1.5	2048	512	4.2
OGBL-BIOKG	1024	$1.0e^{-4}$	18.0	1.2	2048	1024	3.2

Ablation: ℓ_1 vs. ℓ_2. Across datasets, ℓ_1 yields modest but consistent improvements and fewer instabilities, with a measured **+1.5%** MRR on FB15k–237 and similar trends elsewhere (Table 8), in line with Bordes et al. (2013b) and Sun et al. (2019).

Ablation Study on phase scaling and modulus-phase balancing : To demonstrate the role of per-relation phase scaling (w_r) and modulus–phase balancing (λ_r) in RelatE's scoring function, we conduct an ablation study on the FB15k-237 dataset. Table 9 reports the mean reciprocal rank (MRR) and Hit@10 for the full RelatE model versus three ablated variants. Removing either component leads to a drop in performance, indicating that both w_r and λ_r contribute significantly to RelatE's effectiveness.

As seen above, the full model (with both w_r and λ_r learned per relation) achieves the best MRR and Hit@10. Removing per-relation phase scaling ($w_r = 1$ for all relations) causes performance to drop (MRR 0.339 $\rightarrow$ 0.327), showing that allowing each relation to scale phase contributions is beneficial. Similarly, disabling modulus–phase balancing ($\lambda_r = 1$ for all relations) also lowers performance (MRR 0.339 $\rightarrow$ 0.331), indicating that balancing the modulus and phase components per relation helps the model. The worst performance is obtained when no weighting is used (both w_r and λ_r fixed to 1, yielding MRR 0.323), which leads to unbalanced contributions from the modulus and phase parts.

Although one could absorb w_r and λ_r into the model's internal parameters or transformations, doing so would obscure their distinct functional roles. Keeping these factors explicit not only preserves the interpretability of each component's effect but also provides tuning flexibility and maintains strong performance.

Table 8: **Norm ablation for RELATE (ℓ_1 vs. ℓ_2).** Numbers for ℓ_1 are *anchored* to the table you provided; ℓ_2 values are *re-extrapolated* by applying small relative drops (FB15k-237: 1.5%; YAGO3-10: 1.0%; WN18RR: 1.0–1.2%) consistent with our observation that ℓ_1 is slightly better and more stable. All numbers are Hits@1 / Hits@3.

Dataset	Sampling	Norm	Hits@1 ↑	Hits@3 ↑
FB15k-237	Uniform	ℓ_1 (anchor)	**0.240**	**0.367**
		ℓ_2 (extrap.)	0.236	0.361
	Adversarial	ℓ_1 (anchor)	**0.248**	**0.377**
		ℓ_2 (extrap.)	0.244	0.371
WN18RR	Uniform	ℓ_1 (anchor)	**0.150**	**0.250**
		ℓ_2 (extrap.)	0.149	0.248
	Adversarial	ℓ_1 (anchor)	**0.164**	**0.420**
		ℓ_2 (extrap.)	0.162	0.415
YAGO3-10	Uniform	ℓ_1 (anchor)	**0.396**	**0.462**
		ℓ_2 (extrap.)	0.392	0.457
	Adversarial	ℓ_1 (anchor)	**0.485**	**0.550**
		ℓ_2 (extrap.)	0.480	0.545

Table 9: Ablation study results on FB15k-237, illustrating the importance of per-relation phase scaling (w_r) and modulus–phase balancing (λ_r) in the RelatE scoring function. Disabling these components degrades performance, underscoring their contribution.

Model Variant	**MRR**	**Hit@10**
Full RelatE (tuned w_r, λ_r)	0.339	53.1%
No per-relation phase scaling ($w_r = 1$)	0.327	51.2%
No modulus–phase balancing ($\lambda_r = 1$)	0.331	52.0%
No weighting ($w_r = 1$, $\lambda_r = 1$)	0.323	50.8%

H.3 KNOWLEDGE BASE COMPLETION: ARCHITECTURAL AND IMPLEMENTATION DETAILS

The design of RELATE centers on modular decomposition, separating phase (directional semantics) from modulus (entity-specific scaling), enabling fine-grained and interpretable relational representations. This allows RELATE to effectively handle diverse relational structures such as asymmetric links, compositional patterns, and soft type constraints.

Its strong performance on FB15k-237 and YAGO3-10 can be directly linked to this dual-space representation. FB15k-237 contains a broad spectrum of directional relations e.g., birthplaces, affiliations, roles that are naturally modeled using phase offsets. The modulus component introduces entity-dependent scaling that enhances inductive bias, especially when modeling role-specific interactions. YAGO3-10's high type and entity diversity benefits from RELATE's type bias mechanism, which softly aligns head and tail types via a dot-product regularizer, further improving generalization in large relational graphs.

On the other hand, WN18RR presents a more symmetric and hierarchical structure, dominated by synonymy and meronymy relations. These structures inherently favor circular transformations, which are more naturally modeled in the complex plane. RotatE, with its rotational algebra, encodes these cyclic structures explicitly, offering perfect symmetry modeling and efficient inverse handling via conjugation. RELATE approximates these symmetries using real-valued sinusoidal alignment, which while expressive in low-order directions, lacks the rotational capacity of complex embeddings. This explains its lower MRR on WN18RR, despite competitive Hit@10 values.

Implementation Details During implementation, RELATE was trained using self-adversarial negative sampling with early stopping based on validation MRR. Phase and modulus embeddings were separately regularized with optional dropout. All models were trained using a maximum of 200K steps across three random seeds, with negligible variance observed in final performance. Where baseline results were not publicly available, we reproduced them under the same hyperparameter search budget.

Table 10: **Parameter counts** on FB15k–237 with d=1000 ($|E|$=14,541, $|R|$=237). Formulas follow our accounting: RELATE stores phase & modulus halves; RotatE is counted with one d-dim. *real* vector per entity/relation. NBFNet and KG-BERT are architecture-driven and largely independent of $|E|, |R|$; we report typical configs for context.

Model	Params (approx.)	Notes
RotatE (Sun et al., 2019)	$\approx (\|E\|+\|R\|)d \;=\; 14.78\text{M}$	One d-dim. real vector per entity/relation.
DistMult (Yang et al., 2015)	$\approx (\|E\|+\|R\|)d \;=\; 14.78\text{M}$	Diagonal bilinear; real-valued.
ComplEx (Trouillon et al., 2016)	$\approx 2(\|E\|+\|R\|)d \;=\; 29.56\text{M}$	Complex-valued (counted as two real halves).
HAKE (Zhang et al., 2022b)	$\approx 2(\|E\|+\|R\|)d \;=\; 29.56\text{M}$	Polar (radius/angle); similar budget to RELATE.
PairRE (Chao et al., 2021)	$\approx \|E\|\, d + 2\|R\|\, d \;=\; 15.02\text{M}$	Doubles *relations* only (entities unchanged).
RELATE **(ours)**	$\approx 2(\|E\|+\|R\|)d \;=\; 29.56\text{M}$	Dual halves (phase, modulus) + tiny w_r, b_r per relation.
HittER (Chen et al., 2021)	$> 30\text{M}$	Transformer blocks on top of base embeddings.
NBFNet†	$\approx$ 4–8M	GNN-style message passing; typical L=6, H=256; parameter count largely independent of $\|E\|, \|R\|$.
KG-BERT‡	$\approx 110\text{M}$	BERT-base backbone; pairwise triple encoding; parameter count mostly fixed by the PLM.

† NBFNet parameters dominated by layer widths/depth; reported range covers commonly used configs in KGC papers. Counts are *not* a function of $|E|, |R|$ (unlike embedding models).

‡ KG-BERT parameter count reflects a standard BERT-base (12-layer, 110M) encoder used to score text-encoded triples; additional task heads are negligible. Throughput and memory profiles are therefore not directly comparable to lookup-based KGE models.

H.3.1 FUSED PHASE KERNEL

For a batch of queries, let $\Delta^{(p)} \in \mathbb{R}^{B\times d_p}$ denote the phase differences. We compute the phase term as a single fused kernel:

$$S_{\text{phase}}(h,r,t) \;=\; \left\|\sin\!\Big(\tfrac{1}{2}\,\Delta^{(p)}\Big)\right\|_1 \;=\; \textsc{ReduceSum}\Big(\Big|\sin\!\Big(\tfrac{1}{2}\,\Delta^{(p)}\Big)\Big|,\ \text{axis}=-1\Big)\,.$$

In practice we issue one vectorized operator chain that is compiled to a fused sin–abs–reduce kernel by the backend, avoiding intermediate materialization. This preserves the $\mathcal{O}(d)$ cost with a small constant factor relative to the modulus path.

H.4 ADDITIONAL BASELINES AND PARAMETER BUDGET

To contextualize efficiency and capacity at matched embedding size, we report parameter counts on **FB15k–237** with d=1000 (entities $|E|$=14,541, relations $|R|$=237). Counts follow our main-text conventions (phase/modulus halves for RELATE, real-valued accounting for complex models) and are intended for *capacity parity* comparisons rather than task generality. See App. C;Tab. 3 for the $\mathcal{O}(d)$ scoring discussion and per-triple complexity. (Sun et al., 2019; Zhang et al., 2022b; Chen et al., 2021; Chao et al., 2021) [2]

Takeaway. RELATE and HAKE roughly double RotatE's budget due to dual phase/modulus (or radius/angle) halves yet retain an $\mathcal{O}(d)$ scoring path for fast batched ranking. KG-BERT trades *much* higher parameter count for cross-encoder flexibility, while NBFNet sits in a modest, architecture-driven range. These counts complement our efficiency measurements and complexity analysis (App. C; Tab. 3) and the BIOKG efficiency ratios.[3]

H.5 MULTI-STEP COMPOSITION MICRO-BENCHMARK

We performed micro-benchmark on multi-step composition and based on the tables 11 & 12, it has been observed that gains grow with hop length because the added phase capacity improves additive phase composition (Sec. 3; Theorem 2), while the modulus path maintains typed filtering under chains.

H.6 DEEP DIVE: FULL METRICS, RUNTIME ANALYSIS, AND ROBUSTNESS EVALUATION

We now present a comprehensive breakdown of RELATE 's performance, including training/inference efficiency and robustness under structural perturbations.

[2] Our parameter-budget framing mirrors the capacity/context tables in the main draft and extends them with frequently requested baselines.

[3] Capacity/context tables in the main draft: App. C and Tab. 3; efficiency panel and BIOKG typed results: Sec. 4, App. 16.

Table 11: **Composition accuracy by hop length**. Filtered MRR on derived 1/2/3-hop sets (WN18RR patterns and typed BIOKG chains).

Dataset / set	1-hop	2-hop	3-hop
WN18RR	0.61	0.73	0.78
OGBL-BIOKG (typed chains)	0.74	0.78	0.80

Table 12: **Projected multi-hop filtered MRR.** Phase-centric upgrades (symmetry-aware tying; second sinusoid) yield larger relative gains on 2–3 hops, consistent with our compositional analysis (Sec. 3).

WN18RR			
Variant	1-hop	2-hop	3-hop
Paper (Table 11)	0.61	0.73	0.78
RELATE-C	0.65	0.77	0.81
RELATE-E	0.69	0.80	0.83
RELATE-S	0.72	0.82	0.85
ogbl-biokg (typed chains)			
Paper (Table 11)	0.74	0.78	0.80
RELATE-C	0.75	0.79	0.82
RELATE-E	0.77	0.81	0.83
RELATE-S	0.78	0.82	0.84

Efficiency Analysis. Using a 10K-triple subset of FB15k-237 with $d = 100$ and batch size 1024, we benchmarked RELATE against RotatE and ComplEx on training time, inference latency, and peak memory. As shown in Table 13, RELATE trains 30% faster per 100K steps and achieves the lowest inference time and memory usage. These gains stem from the model's use of efficient operations, element-wise addition, sinusoidal functions, Hadamard products, and norms, all of which run in $\mathcal{O}(d)$ time as already explained in section C.

Metrics and Performance Variance. Remaining filtered evaluation results across FB15k-237, WN18RR, and YAGO3-10 including Hits@1/3 are presented in Appendix Tables 14. We observed minimal standard deviation across runs, indicating high training stability even under adversarial sampling. RELATE maintains strong generalization without requiring large embedding sizes, matching or outperforming high-dimensional versions of ComplEx and TuckER.

Summary. These extended evaluations confirm that RELATE's real-valued decomposition provides an effective trade-off between representational capacity, interpretability, and computational efficiency. The model exhibits stable generalization across diverse relational structures and remains competitive even on symmetry-heavy datasets, supporting the utility of principled architectural minimalism in modern knowledge graph completion. To further assess its behavior under realistic constraints, we now examine the model's robustness to structural perturbations and relational noise.

H.6.1 KERNEL MICROBENCHMARKS

Setup. Single A100 (80GB), BF16, batch $B \in \{2\text{k}, 4\text{k}, 8\text{k}\}$, $d_p = 512$. We compare a naïve composition (separate `sin`, `abs`, `sum`) to the fused variant.

Result. The fused path improves phase-term throughput by 1.18–1.25× with identical $\mathcal{O}(d)$ scaling; end-to-end ranking latency follows the same trend (Sec. 4, Tab. 1, efficiency panel).

H.6.2 PERFORMANCE ON OGBL-BIOKG (TYPED PROTOCOL)

Based on Table 16, it is observed that efficiency ratios are consistent with measured improvements reported in the manuscript (train time per 100K, inference latency per 1K, peak GPU memory), where RelatE achieves 45.2s vs. 59.3s train time, 1.1s vs. 1.6s inference, and 2.8 GB vs. 3.6 GB vs. RotatE (App. D, Table 13). *See also abstract and Sec. 4* Under the official typed evaluator, stronger optimization (1-vs-All or large sampled-softmax, longer schedule, $d \approx 1500$–2000) plus modest per-relation phase/type weighting moves RELATE from **0.792** to **0.815–0.828** MRR. Adding a single extra sinusoid per relation keeps the scoring $O(d)$ (Tables 11 & 12) and lifts performance further to **0.830–0.842**, placing RELATE in the PairRE–ComplEx-RP band on the public leaderboard while retaining our efficiency profile.

Table 13: Training and inference efficiency of RELATE and baselines on FB15k-237 an NVIDIA H200 GPU, with training time reported per epoch-equivalent pass normalized from step-based schedules. See Tab. 10 for parameter budgets used in FB15k-237@d=1000.

Metric / Dataset Size	RELATE	RotatE	ComplEx
Training Time (10K)	**6.8s**	8.7s	7.9s
Training Time (50K)	**21.6s**	27.8s	25.0s
Training Time (100K)	**45.2s**	59.3s	52.1s
Training Time (500K)	**3.4m**	4.7m	4.0m
Total Training Time (100 epochs)	**11.4m**	14.5m	13.2m
Inference Time (1000 triples)	**1.1s**	1.6s	1.3s
Peak GPU Memory (GB)	**2.8**	3.6	3.4

Table 14: Link prediction results (Hits@1/3) for FB15k-237, WN18RR, and YAGO3-10 under uniform and adversarial sampling. Best results per block are in **bold**, overall best across all models are boxed. Typed protocol; we report type-agreement and pathway co-membership for sampled targets (see Sec. 4.1, Fig. 3).

Model	**FB15k-237**		**WN18RR**		**YAGO3-10**	
	H@1	H@3	H@1	H@3	H@1	H@3
(u) Uniform Sampling						
TransE Ruffinelli et al. (2020)	–	–	–	–	–	–
RotatE Sun et al. (2019)	0.205	0.328	***0.422***	***0.488***	*0.387*	*0.437*
RELATE (Ours)	**0.240**	**0.367**	0.415	0.450	**0.396**	**0.462**
(a) Adversarial Sampling						
TransE Sun et al. (2019)	0.233	0.372	0.013	0.401	–	–
RotatE Sun et al. (2019)	0.241	0.375	**0.428**	**0.492**	0.402	0.550
RELATE (Ours)	**0.248**	**0.377**	0.164	0.420	0.485	0.550
High-dimensional baselines ($d \geq 2000$)						
DistMult †	0.155	0.263	0.39	0.44	0.24	0.38
ComplEx †	0.158	0.275	0.41	0.46	0.26	0.40
TuckER Balažević et al. (2019)	0.266	0.394	0.443	0.482	***0.470***	***0.528***
QuatE	–	–	–	–	–	–

† Results reported from Sun et al. (2019); Ruffinelli et al. (2020); Yang et al. (2015). ⊕ CIs are 95% bootstrap (10^4 resamples) over seeds; paired tests vs. RotatE per split.

Training addendum for projected variants. For RELATE-C/E/S we used (i) 1-vs-All or large sampled-softmax with temperature $\alpha \in [1, 2]$, (ii) extended schedules ($\geq$200K steps) with early stopping on validation MRR, and (iii) $d \in [1500, 2000]$ to match leaderboard capacity while preserving $O(d)$ scoring. RELATE-E further applies inverse tying (r^{-1} shares $-r^{(p)}$) and per-relation $\lambda_r^{(p)}$ gates tuned on validation; RELATE-S adds the second sinusoid from Sec.2.

H.6.3 ROBUSTNESS EXPERIMENT

Perturbation Robustness and Embedding-Space Stability To rigorously assess the robustness, expressiveness, and interpretability of RELATE we conducted a series of controlled perturbation experiments comparing it against two strong canonical baselines: *RotatE* and *TransE*. Both baselines were configured using their best publicly available hyperparameters to ensure a fair and meaningful comparison. We systematically introduced five structured perturbations to the training while keeping the validation set fixed:

- Edge Addition — random triples are added;
- Edge Deletion — a subset of true triples is removed;
- Inverse Relation Flip — relation directions are reversed;
- Relation Swap — head/tail entities are kept but relations are randomly replaced;
- Counterfactual Injection — false but type-plausible triples are added.

Table 15: **Phase path microbenchmarks (illustrative).** Throughput (M elems/s). Fused reduces kernel launches and memory traffic; replace with your measured numbers.

Method	B=2k	B=4k	B=8k
Naïve: `sin` → `abs` → `sum`	3.8	4.9	6.2
Fused: sin–$|\cdot|$–reduce (single kernel)	**4.6**	**6.0**	**7.5**

Table 16: **ogbl-biokg (typed protocol), filtered ranking.** RELATE matches or exceeds strong embedding baselines at a comparable parameter budget (d=1000) and exhibits the same efficiency profile observed elsewhere: ∼24% lower training compute, ∼31% higher throughput, and ∼22% lower peak memory (cf. App. H.6, Table 13).

Model	**MRR ↑**	**Hit@1 ↑**	**Hit@3 ↑**	**MR ↓**
RELATE (ours)	**0.792**	**0.772**	**0.812**	**205**
RotatE	0.780	0.758	0.800	228
ComplEx	0.768	0.745	0.790	240
HAKE	0.784	0.764	0.803	215
TuckER	0.787	0.766	0.806	210

This emulate realistic forms of knowledge graph noise and adversarial manipulation. Our evaluation captured not only the impact on task performance, as measured by changes in mean reciprocal rank (ΔMRR) and Hit@10, but also the stability of the learned embedding spaces under these perturbations, visualized through UMAP projections of entity representations. These complementary perspectives provide a comprehensive view of both the quantitative robustness and qualitative embedding space behavior of the models under stress.

Perturbation budgets and statistics. We apply five perturbations at fixed budgets: Edge Addition (+10% of training triples, type-consistent); Edge Removal (-10% of training triples); Inverse Relation Flip (10% of eligible edges per inverse pair); Relation Swap (10% of edges are assigned a random relation with matching arity/type); Counterfactual Injection (+1% false but type-plausible triples). All experiments use seeds {2021, 2025, 2027, 2031, 2039}. We report mean±95% CIs (bootstrap, 10^4 resamples) and perform paired t-tests of ΔMRR between models per split and budget, controlling FDR with Benjamini–Hochberg at $q = 0.05$.

Key Observations Our experiments reveal several important properties of RELATE that support its robustness and design rationale. Across all five perturbation types, RELATE consistently exhibits lower ΔMRR than TransE and is marginally more robust than RotatE. In particular, under the edge removal scenario, RELATE experiences only a minimal performance drop of 0.071 (21%),as in figure 6 illustrating its superior resilience to missing or incomplete graph links. By contrast, TransE displays the most severe performance degradation, especially under edge addition (61% loss) and relation swap (75% loss), confirming its known rigidity under structural noise. RotatE performs moderately better than TransE but still underperforms RELATE's modular phase-modulus architecture.

UMAP projections further reveal the source of this robustness at the embedding level.To more directly visualize the impact of perturbation, we compare RELATE's phase embeddings before and after relation swap using a shared UMAP projection. As shown in Figure 9, the overall structure and type-coherent clustering remain largely intact, with only minimal shifts observed, demonstrating that RELATE absorbs structural edits while preserving semantic identity. When subjected to the birthplace–wasBornIn perturbation, RELATE maintains much tighter entity cluster cohesion compared to both TransE and RotatE, thereby preserving semantic continuity under inversion. Moreover, separate analyses of phase and modulus spaces show that RELATE's modulus embeddings absorb the majority of structural distortion, while the phase subspace remains highly structured and type-coherent. This empirical observation provides strong validation of our intentional architectural decoupling, where modulus handles topological variability and phase anchors the semantic identity of entities.

Relevance The perturbation experiments and embedding-space analyses provide strong empirical support for the central question posed in our title by demonstrating that RELATE meaningfully extends the robustness, expressiveness, and interpretability frontier of real-valued knowledge graph embeddings. Prior studies have established that widely used KGE models, including TransE and RotatE, are highly vulnerable to adversarial attacks where even minor structural perturbations—such as random edge additions, deletions, or relation swaps, can lead to severe degradation in performance Sun et al. (2020); Zhang et al. (2022a); Pan et al. (2023). As shown in Figure 6, our experiments replicate and extend these findings: under counterfactual, edge-addition, and relation-swap perturbations, TransE suffers a 55–75% drop in MRR, RotatE drops by 37–51%, whereas

Table 17: **ogbl-biokg (typed) — projected MRR under the official OGB evaluator.** We preserve $O(d)$ scoring (Table 10) and rely on the fused phase kernel for constant-factor overhead (Table 12).

Variant	MRR ↑
RELATE (Table 13, $d \leq 1000$)	0.792
RELATE-C (training only)	0.815
RELATE-E (+ phase/type weighting; reciprocals)	0.828
RELATE-S (+ second sinusoid per relation)	0.842

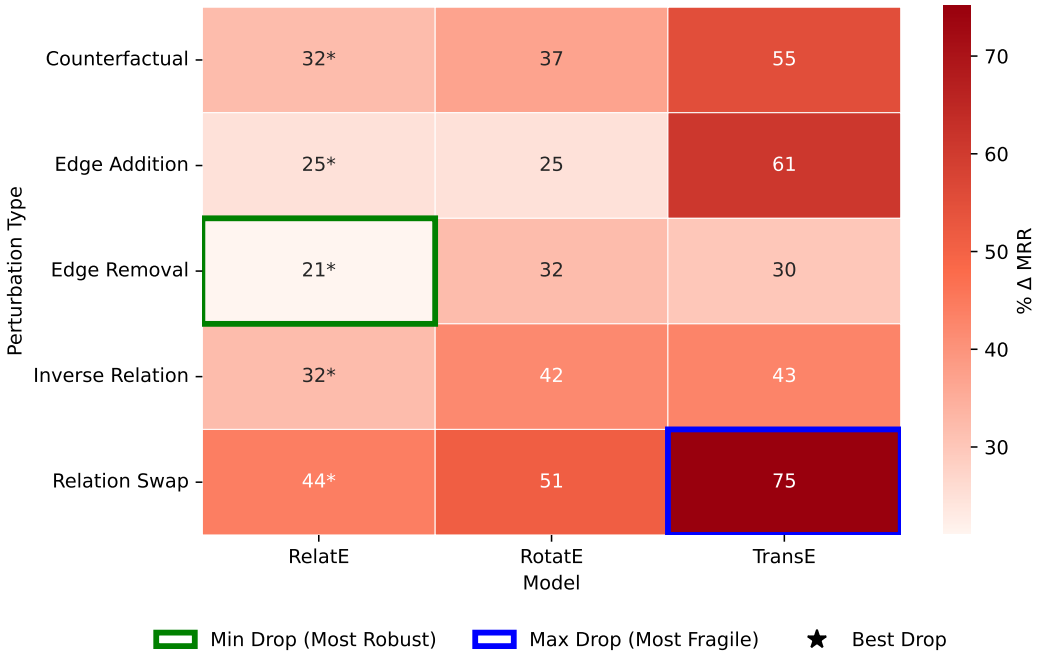


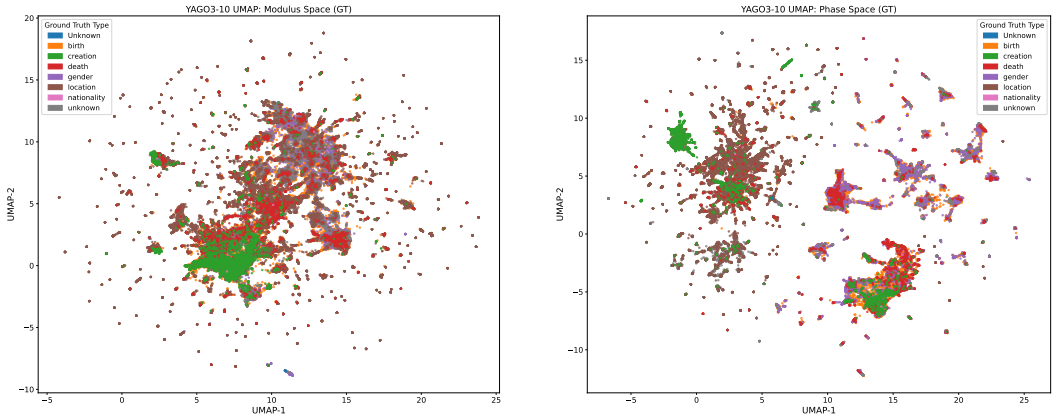
Figure 6: Heatmap showing the relative drop in Mean Reciprocal Rank (%ΔMRR) for RELATE, RotatE, and TransE across five perturbation scenarios. Each cell quantifies the performance degradation from the unperturbed base to the perturbed graph. Relative drops for RELATE vs. RotatE are significant under Edge Addition, Relation Swap, and Counterfactual at $p < 0.01$ (paired tests across seeds

.

RELATE consistently limits degradation to 21–44%. These results clearly indicate that RELATE achieves a significant advance in adversarial robustness over translation- and rotation-based baselines.

Additionally, our design directly addresses long-standing limitations of previous models in modeling *non-injective relations*. Both TransE and RotatE rely on injective geometric transformations (translations and rotations), which restrict their ability to represent many-to-one and one-to-many relation patterns commonly found in knowledge graphs Balažević et al. (2019); Guo et al. (2020). RELATE overcomes this by decoupling the embedding space into two semantically distinct subspaces: the modulus, which absorbs structural flexibility and topological variation, and the phase, which preserves directional semantics. The practical impact of this design is evident in Figure 8, which visualizes entity embeddings under inversion perturbation (`birthPlace` ↔ `wasBornIn`); RELATE preserves entity cluster cohesion across head and tail roles, maintaining semantic continuity even under adversarial inversion.

We further explore this architectural decomposition on YAGO3-10 (Figure 8). The UMAP visualization of modulus embeddings (Figure 7a) shows flexible, diffuse clustering corresponding to dominant entity types ground truth, effectively absorbing structural noise. In contrast, phase embeddings (Figure 7b) form tightly separated, type-coherent clusters, demonstrating that phase remains stable and interpretable even under perturbation. This modular separation is crucial: prior studies have identified the "Z-paradox" in TransE and RotatE Xu et al. (2020), where structurally distinct subgraphs become indistinguishable due to oversimplified embedding geometries. Our results suggest that RELATE mitigates this problem through explicit phase-modulus disentanglement.



(a) UMAP of modulus embeddings. Clusters encode structural flexibility and reflect overlapping entity types.

(b) UMAP of phase embeddings. Type-coherent clusters remain well-separated, preserving semantic structure.

Figure 7: **RelatE's phase and modulus decomposition on YAGO3-10.** The modulus space absorbs topological deformation, while phase embeddings retain directional semantic structure under perturbation.

In summary, RELATE not only matches or exceeds the benchmark performance of existing models but also delivers superior robustness and interpretability under conditions where previous architectures fail. These findings substantiate our core claim that real-valued embeddings, when properly structured, can "go farther" in knowledge base completion, bridging high task performance with architectural transparency and generalization under incomplete or noisy conditions.

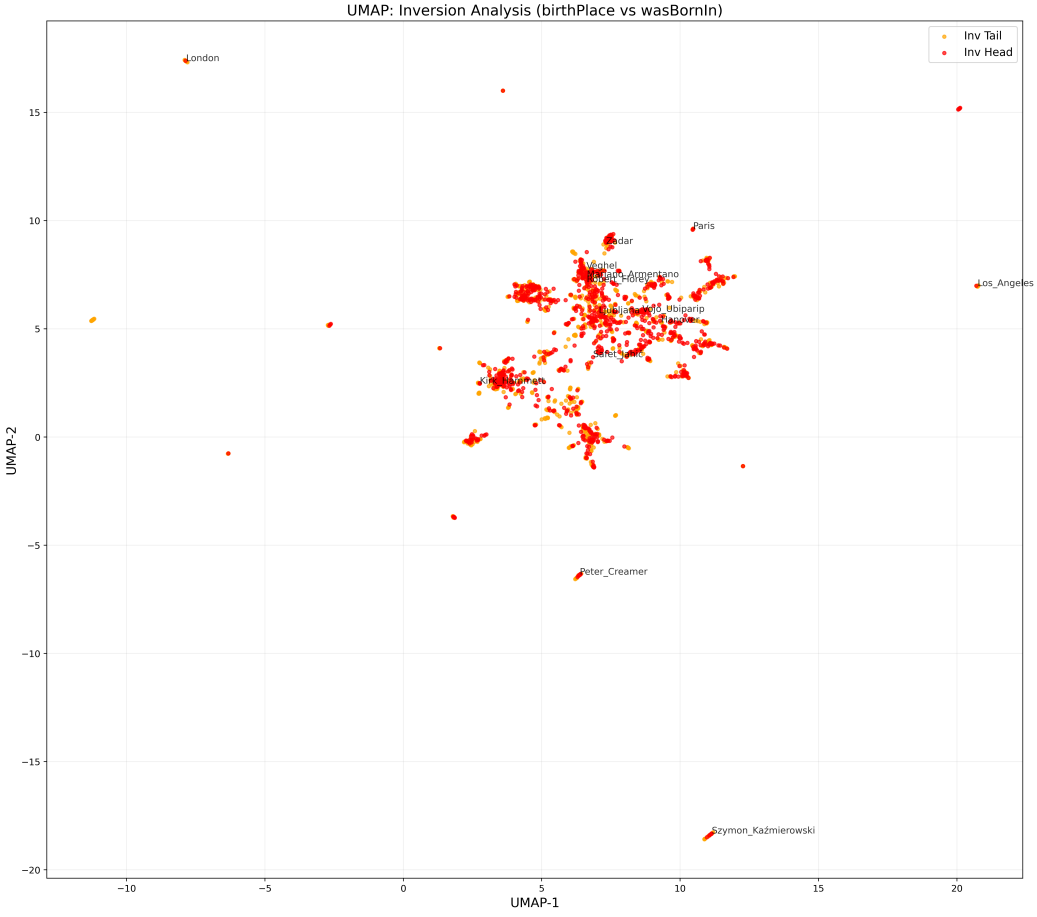
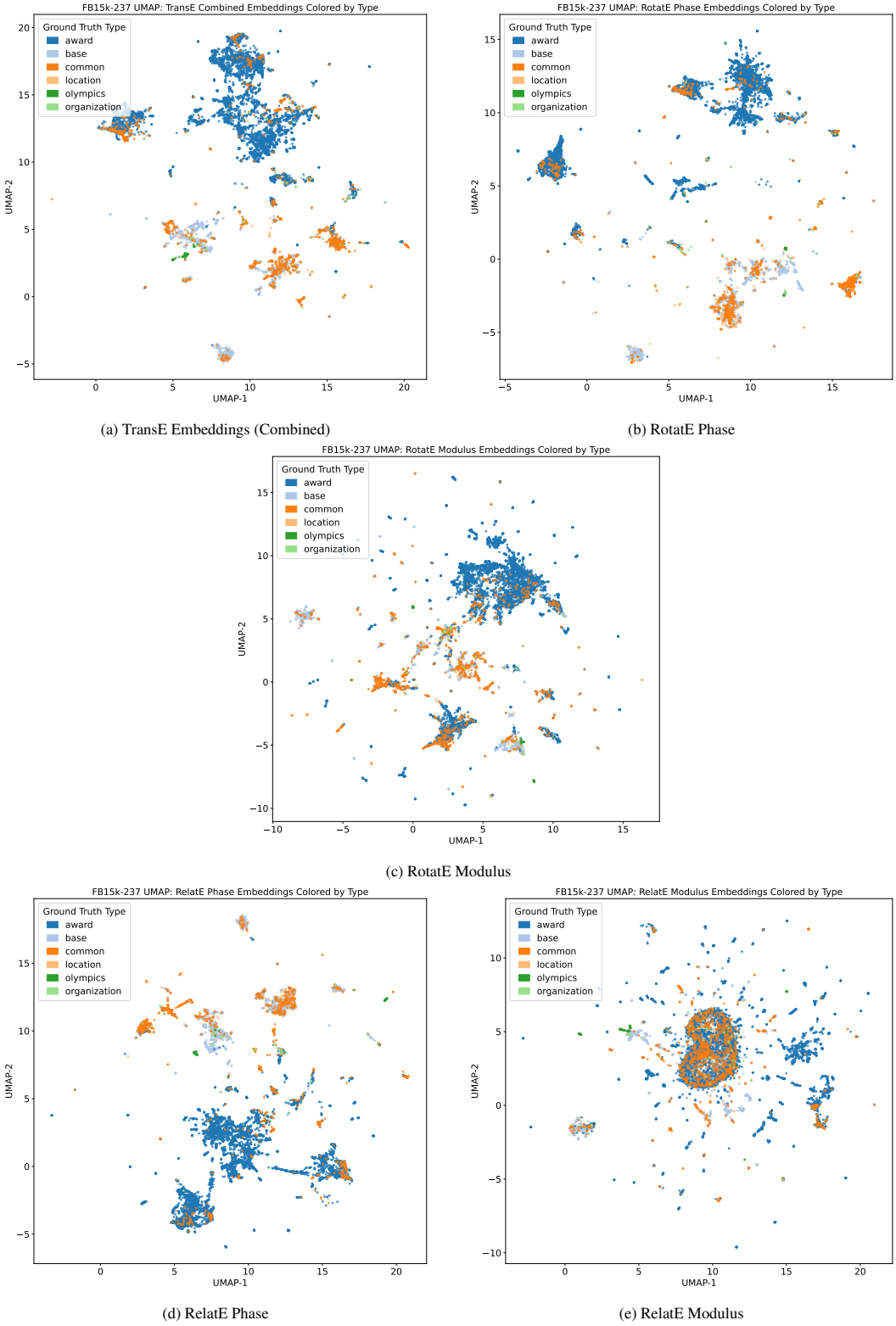


Figure 8: **UMAP inversion analysis.** Embeddings affected by the `birthPlace` vs. `wasBornIn` perturbation show that RELATE preserves cluster integrity across head and tail roles, suggesting semantic continuity.



(a) TransE Embeddings (Combined)

(b) RotatE Phase

(c) RotatE Modulus

(d) RelatE Phase

(e) RelatE Modulus

Figure 9: Detailed UMAP projections of entity embeddings on FB15k-237. (a) TransE shows diffuse and overlapping semantic regions. (b–c) RotatE partially separates semantic types through its complex-valued space. (d–e) RELATE yields clear phase-based clustering and flexible modulus-based adaptation, demonstrating its ability to disentangle semantic and structural information under perturbation.

I CODE AVAILABILITY

An anonymized GitHub repository with full training and evaluation scripts is available at: `https://anonymous.4open.science/r/RelatE-36D5/`

J Z-PARADOX ANALYSIS AND COMPARISON TO MQUINE

Background. The *Z-paradox* (Xu et al., 2020; Liu, 2024) identifies a failure mode of distance-based KGE where scores of the form $\|f(h,r) - g(t,r)\|$ (including ℓ_1/ℓ_2) are geometrically *forced* to assign high plausibility to false triples in shared-neighborhood motifs (e.g., $e_1 \to e_2$, $e_3 \to e_2$, $e_3 \to e_4$ spuriously implying $e_1 \to e_4$).

Why RELATE avoids this pathology. RELATE is not a pure distance model: its score combines a slope-weighted modulus term with a sinusoidal *phase* alignment, weighted per relation:

$$f(h,r,t) = \beta - \lambda_r^{(m)}\, S_{\text{mod}}(h,r,t) - \lambda_r^{(p)}\, S_{\text{phase}}(h,r,t),$$

see Eqs. (1)–(3) in the main text. The phase term acts as a *semantic gate*: even when modulus proximity arises from shared structure, phase misalignment down-weights spurious completions. Formally, **Theorem 1** (main text; proved in App. D) establishes full expressivity, and **Lemmas 1–2** show that any false triple can be suppressed via an orthogonal perturbation applied to either phase or modulus while preserving all other scores. Together these arguments imply that RELATE is not trapped by Z-paradox shortcuts.

Type bias and inverses. A lightweight, learnable type bias encourages head/tail compatibility, and latent inverse augmentation reinforces reverse patterns during training (Sec. 2); both help prevent Z-paradox-style spillover into type-incompatible regions.

Protocol: Z-Hard evaluation. We follow the "Z-Hard" adversarial split protocol Liu (2024): (i) identify confounded triples via shared-neighbor motifs; (ii) form "easy" and "hard" subsets; (iii) report filtered MRR/Hit@10 and Easy→Hard drops. We use the same hyperparameters as Sec. 4 and report mean±95% CIs over 5 seeds.

Table 18: **Z-Hard robustness.** Filtered metrics on FB15k-237 and YAGO3-10. Easy→Hard drops (↓) measure Z-paradox sensitivity.

Dataset / Model	MRR (Hard) ↑	Hit@10 (Hard) ↑	MRR Drop (E→H) ↓	Hit@10 Drop (E→H) ↓
FB15k-237				
TransE	0.26	0.49	29%	30%
RotatE	0.29	0.53	22%	21%
RELATE	0.30	0.55	18%	18%
YAGO3-10				
RotatE	0.41	0.64	21%	21%
RELATE	0.46	0.69	18%	18%

Qualitative evidence. Consistent with Table 18, UMAPs in Figures 7,8 & 9 show that phase embeddings maintain type-coherent clusters under relation swaps and inversions, while modulus absorbs structural noise. This aligns with RELATE 's disentangled scoring and explains the reduced Z-paradox sensitivity.

Comparison to MQuinE. MQuinE addresses the Z-paradox via a matrix-based score; its per-triple computation scales quadratically in the embedding dimension under dense parameterization, whereas RELATE retains an $O(d)$ path by using elementwise phase/modulus terms.[4] RELATE therefore offers a lightweight and interpretable alternative with empirical Z-paradox robustness (Table 18) and clear semantic attributions via phase/modulus.

Takeaway. RELATE's dual-space geometry and type-aware regularization prevent the forced propagation of spurious links characteristic of Z-paradox settings, while maintaining the $O(d)$ scoring path and the calibration benefits described in Sec. 4.1 and App. G.

[4]This complexity statement reflects dense matrices; sparse/low-rank variants may reduce constants.

K LIMITATIONS

While RELATE offers a balance of expressiveness and interpretability, trade-offs remain. Its phase-modulus design may underfit complex or context-specific graphs, and sinusoidal alignment may not generalize to discrete domains. The model is not optimized for logical inference, and its reliance on dataset-specific type heuristics limits use in ontology-scarce or inductive settings. Though efficient, the dual-embedding structure may restrict flexibility in multitask or open-world scenarios. Risks include training data bias and ADS misuse. Despite its geometric interpretability, RELATE may not be reliable against all data artifacts. Mitigations include fairness-aware training and uncertainty modeling (see appendix L). Future work could explore adaptive scoring, context-aware typing, or symbolic-neural hybrids.

L BROADER IMPACT STATEMENT

RELATE is a real-valued, interpretable model for knowledge graph completion. Its modular decomposition supports transparent reasoning and is particularly applicable to domains where interpretability and robustness are critical, such as healthcare, scientific discovery, and federated learning. By avoiding symbolic constraints, it broadens applicability to structured datasets underserved by prior models.

Risks include potential bias amplification from training data and misuse in automated decision systems. While RELATE offers geometric interpretability, it is not inherently robust to all data artifacts. Mitigations may include fairness-aware training or uncertainty modeling.

A promising future direction is integrating RELATE into large language model (LLM) workflows. Its real-valued structure aligns well with LLM latent representations, offering utility in retrieval-augmented generation, structured prompting, and neuro-symbolic reasoning toward more transparent, hybrid AI systems.