Double Permutation Equivariance for Knowledge Graph Completion

Anonymous Authors[1]

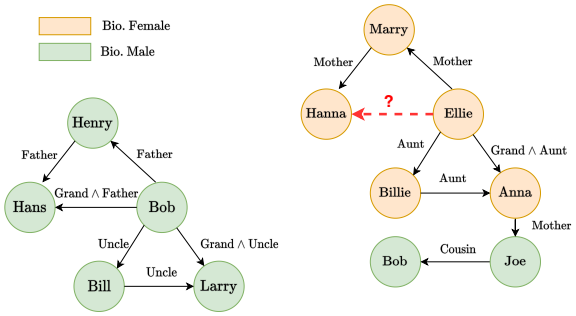
Abstract

This work provides a formalization of Knowledge Graphs (KGs) as a new class of graphs that we denote *doubly exchangeable attributed graphs*, where node and pairwise (joint 2-node) representations must be equivariant to permutations of both node ids and edge (& node) attributes (relations & node features). Double-permutation equivariant KG representations open a new research direction in KGs. We show that this equivariance imposes a structural representation of relations that allows neural networks to perform complex logical reasoning tasks in KGs. Finally, we introduce a general blueprint for such equivariant representations and test a simple GNN-based double-permutation equivariant neural architecture that achieve 100% Hits@10 test accuracy in both the WN18RRv1 and NELL995v1 inductive KG completion tasks, and can accurately perform logical reasoning tasks that no existing methods can perform, to the best of our knowledge.

1. Introduction

Knowledge graphs (KGs) are generally defined as structured representations of collections of facts in the form of a set of triplets $\mathcal{S} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$, where $(i, r, j) \in \mathcal{S}$ define two entities i (head entity) and j (tail entity) connected by a relation r, where both nodes and relations are finite: $N = |\mathcal{V}| < \infty$ and $R = |\mathcal{R}| < \infty$. In some applications KGs naturally define conjunctive logical statements (as in Figure 1(a)): $(i, \text{Father}, j) \wedge (j, \text{Father}, u) \wedge (i, \text{Grand}, u) \wedge (i, \text{Father}, u) \wedge$ etc., where $\mathcal{R} = \{\text{Father}, \text{Grand}, \ldots\}$ and $\mathcal{V} = \{i, j, u, \ldots\}$.

Unfortunately, KGs are often incomplete. Hence, the task of predicting missing relations (e.g., predict missing $(i, r, j) \in \mathcal{V} \times \mathcal{R} \times \mathcal{V}$) is both widely-studied and a key task in knowledge base construction, often denoted as *knowledge graph completion* (Bordes et al., 2013; Nickel et al., 2015; Teru et al., 2020). As defined above (and in the literature (Lin et al., 2015; Chen et al., 2020b; Kejriwal et al., 2021; Shen et al., 2022)), knowledge graph completion is the task of predicting attributed edges, where not only we need to identify that a pair of nodes $(i, j) \in \mathcal{V} \times \mathcal{V}$ is a missing edge in the KG, but also determine which relation $r \in \mathcal{R}$ the edge (i, j) has.



(a) Training KG

(b)Test KG (new nodes & relations)

Figure 1. (Biological human KG) Illustrative knowledge graphs of biological human relations. Exemplar inductive task: learn on training KG (a) to inductively predict missing relation "?" over Test KG (b) with new nodes (potentially more), new relations (potentially more), and new node features (potentially more).

Treating KGs as attributed graphs allows researchers to adapt Graph Neural Network (GNN) methods used for link prediction with only minor modifications: Distinct pooling operations for each edge type (regularized to avoid overfitting), and changing the output from binary classification (edge prediction) to multi-label classification (predicting R relation labels). Roughly, this is the formula followed by RGCN (Schlichtkrull et al., 2018), GraIL (Teru et al., 2020), NodePiece (Galkin et al., 2021), NBFNet (Zhu et al., 2021), and ReFactorGNNs (Chen et al., 2022), among others. *However, theoretically, are KGs just attributed graphs?*

Contributions. *In this work we argue that some KGs belong to a new class of graphs (which we denote as* doubly exchangeable attributed graphs*) whose node and pairwise representations must be equivariant to the action of the permutation group composed by the permutation subgroups of node ids, edge attributes (relations), and node*

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

attributes (theoretically modeled as self-edge attributes without loss of generality). This equivariance imposes a type of structural learning akin to inductively learning to answer a subset of Horn clauses in the test KG where both entities and relations are subject to universal quantifiers (Definition 4.9), e.g., for a given pair $i, j \in \mathcal{V}^{\text{test}}$, and $\forall r_1, r_2 \in \mathcal{R}^{\text{test}}, \forall u \in \mathcal{V}^{\text{test}}, (i, r_1, u) \wedge (u, r_2, j) \implies (i, r_2, j)$. If $\mathcal{S}_1^{\text{train}}, \mathcal{S}_2^{\text{train}}, \ldots \subseteq \mathcal{V}^{\text{train}} \times \mathcal{R}^{\text{train}} \times \mathcal{V}^{\text{train}}$ are the training KGs, this equivariance allows the trained predictor to perform predictions over $\mathcal{S}^{\text{test}} \subseteq \mathcal{V}^{\text{test}} \times \mathcal{R}^{\text{test}} \times \mathcal{V}^{\text{test}}$ in the test KG, where $\mathcal{V}^{\text{test}}$ and $\mathcal{V}^{\text{train}}$, $\mathcal{R}^{\text{test}}$ and $\mathcal{R}^{\text{train}}$ are all potentially distinct sets (with potentially distinct sizes).

We believe our double-permutation equivariance could be similarly impactful for KG machine learning as (single) permutation-equivariance has been for graph machine learning. Our work will focus on inductive KG completion tasks, but the same methods can also be applied transductively. There is no existing KG completion task in the literature that can test the full power of our approach. Our experimental results use the inductive WN18RR-v1 and NELL995-v1 KG completion tasks, which have new nodes in test but not more and completely new relations. We also design two harder synthetic tasks. No existing KG completion method is able to perform one of our synthetic tasks (see Table 2), where the test KG has more and new relations than the training KG.

2. Theory review: What are attributed graphs? An exchangeability perspective

Theoretically, a multigraph —we will refer to multigraphs as graphs— with $N \geq 2$ vertices is a sequence of edges $\mathbf{A}^{(N,R)} = (A_{1,1}, A_{1,2}, \ldots, A_{N,N}) \in \mathbb{A}_R^{N^2}$, where $\mathbb{A}_R$ is some arbitrary domain that encodes $R \geq 1$ relation attributes —e.g., in KGs $\mathbb{A}_R$ is a vector representing multiple edges and their attributes. Without loss of generality, node attributes will be defined as a special type of edge reserved for self-loops $A_{i,i}$. In most applications, what distinguishes a graph from a sequence is the assumption that the choice of node ids to create this sequence is arbitrary. Hence, any prediction that uses $(A_{1,1}, \ldots, A_{N,N})$ as input should be invariant[1] to the permutation of node ids (Srinivasan & Ribeiro, 2020). In statistics, this property is known as joint (array) exchangeability (Aldous, 1981). GNNs (without positional encoding) are permutation-equivariant representation functions, possessing the correct invariances for node and graph-wide classification tasks (Xu et al., 2019a; Morris et al., 2019; Srinivasan & Ribeiro, 2020). Link prediction is better served by equivariant pairwise representations (Srinivasan & Ribeiro, 2020).

More precisely, and without loss of generality, let $\mathcal{V}^{(N)} = \{1, \ldots, N\}$ be the set of nodes (i.e., node ids). For consistency of the notation with knowledge graph, we denote $(\mathbf{A}^{(N,R)})_{i,r,j} = A_{i,j,r}$. Let $\pi \in \mathbb{S}_N$ be a permutation from the symmetric group $\mathbb{S}_N$ with degree N, and $(\pi \circ \mathbf{A}^{(N,R)})_{\pi \circ i, r, \pi \circ j} = (\mathbf{A}^{(N,R)})_{i,r,j}$ be the action of permutation π on the sequence $(A_{1,1}, \ldots, A_{N,N})$, which permutes node ids according to π, that is, $\pi \circ i = \pi_i$, $\forall (i, r, j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$. A function that outputs a d-dimensional node representations of any-size graphs is defined as $\Gamma_{\text{node}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} (\mathcal{V}^{(N)} \times \mathbb{A}_R^{N^2}) \to \mathbb{R}^d$, $d \geq 1$, should be invariant to node id permutations. That is, $\Gamma_{\text{node}}(i, \mathbf{A}^{(N,R)}) = \Gamma_{\text{node}}(\pi \circ i, \pi \circ \mathbf{A}^{(N,R)}), \forall i \in \mathcal{V}$.

Similarly, a neural network that outputs d-dimensional pairwise representations of any-size graphs is defined as $\Gamma_{\text{pair}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} (\mathcal{V}^{(N)} \times \mathcal{V}^{(N)} \times \mathbb{A}_R^{N^2}) \to \mathbb{R}^d$. Pairwise representation should also be invariant to the action of any $\pi \in \mathbb{S}_N$, i.e., $\Gamma_{\text{pair}}((i, j), \mathbf{A}^{(N,R)}) = \Gamma_{\text{pair}}((\pi \circ i, \pi \circ j), \pi \circ \mathbf{A}^{(N,R)})$.

We can also define a graph-wide representation $\Gamma_{\text{gra}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2} \to \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{R}^{N \times R \times N \times d}$, noting that only the mappings between the domain and image that have the same values of R and N are possible. The representation Γ_{gra} is equivariant, that is, for $\pi \in \mathbb{S}_N$, $\pi \circ \Gamma_{\text{gra}}(\mathbf{A}^{(N,R)}) = \Gamma_{\text{gra}}(\pi \circ \mathbf{A}^{(N,R)})$.

GNNs and nearly all recent advances in graph representation learning are driven by the above invariances and equivariances (Bronstein et al., 2017; Chen et al., 2020a; Defferrard et al., 2016; Gilmer et al., 2017; Gori et al., 2005; Hamilton et al., 2017; Maron et al., 2019; Morris et al., 2019; Murphy et al., 2019b;a; Srinivasan & Ribeiro, 2020; Teru et al., 2020; Xu et al., 2019a; Zhu et al., 2021).

3. Brief Related Work: Knowledge graphs as attributed multigraphs

Should there be extra assumptions in some KGs beyond the joint node id permutation exchangeability of attributed graphs? In an excellent introduction on KGs, Kejriwal et al. (2021) warns the reader that "however, we must also deal with the uncomfortable notion that KG is still not very well defined (which makes KG representation challenging because no one representation can be held to be "correct")." In what follows we provide a brief history of knowledge graphs in the literature.

To the best of our knowledge the term *knowledge graph* was first introduced by Schneider (1973) to describe a tutoring system, where each node describes a concept and each arc (direct edge) describes an attributed association between concepts. By 2012, KGs received renewed interest when Google revealed them as a key ingredient in its successful search engine, "things not strings" as described in Singhal (2012). In light of recent advances in large language models (Schulman et al., 2022), the discussion whether

[1] Refer to the theory in (Srinivasan & Ribeiro, 2020) for the sufficiency of invariances.

knowledge can be described by things or strings gain renewed interest. And we believe our work sheds new light into this discussion.

The view of KGs as attributed graphs was somewhat consolidated in the *semantic web* literature around 2016 (Kroetsch & Weikum, 2016; Paulheim, 2017) and by early work on knowledge bases (Bordes et al., 2011), that later was able to integrate classical AI methods (based on knowledge bases and logic), statistical relational learning (SRL) (De Raedt, 2008; Koller et al., 2007; Kersting & De Raedt, 2007; Heckerman et al., 2007; Neville & Jensen, 2007), and attributed graph completion methods KGs (Bordes et al., 2013; Nickel et al., 2015; Teru et al., 2020).

In the SRL literature (e.g., Raedt et al. (2016)) the attribute of an edge $(i, j) \in \mathcal{V} \times \mathcal{V}$ is sometimes instantiated as either a node $r \in \mathcal{R}$ or a node $r(i, j)$, where r is the edge attribute (relation) (e.g., Heckerman et al. (2007)). The drawback of adding edge attributes (relations in $\mathcal{R}$) as nodes in a Bayesian network is that Bayesian networks are sequences (non-exchangeable), but, if treated as a graph (exchangeable), nodes and relations would be exchangeable among themselves (which in many KG applications would be incorrect, since $\mathcal{V}$ and $\mathcal{R}$ are fundamentally distinct sets). Exchangeability w.r.t. node ids in SRL appears in the form of lifting for parameterized templated graphical models, see Koller & Friedman (2009) and Raedt et al. (2016, Chapter 3.1). In practice, automatically finding these templates is difficult and tends to underperform when compared to more modern attributed graph methods for KG completion.

State-of-the-art methods for KG completion treat KGs as attributed (multi)graphs (i.e., only node id exchangeable). They include tensor factorization methods (Bordes et al., 2013; Trouillon et al., 2016; 2017; Sun et al., 2019) (mostly applied in transductive KG tasks) and graph neural network methods (GNNs) (Chen et al., 2022; Schlichtkrull et al., 2018; Galkin et al., 2021; Teru et al., 2020; Wang et al., 2021; Zhu et al., 2021) (mostly applied in inductive KG tasks), among others. Interestingly, out of those, the most successful embedding methods (tensor or GNNs) tend to impose some form of TransE-style (Bordes et al., 2013) translation equivariance in the embeddings (or impose rotation invariance). This embedding equivariance is *markedly different* from relation equivariance, since here each relation has its own personalized shift. Due to space constraints, a more detailed discussion of related work can be found in Appendix B.

4. Proposal: Define some KGs as double exchangeable attributed (multi)graphs

In the following text, we provide definitions and theoretical statements of our proposal in the main paper, while referring all proofs to Appendix A. Our model is intended for a broad class of KGs (but not all KGs may satisfy our conditions). The proposal starts with defining the concept of knowledge graph used in this paper:

Definition 4.1 (Knowledge Graph (KG))**.** A knowledge graph is a multigraph $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$ sampled as $\mathbf{A} \sim \mu$, where μ is some unknown distribution, and $\mathbb{A}_R$ is the set encoding edge attributes (relations) [2]. For instance, if $\mathbf{A}$ has no node attributes, we can define $\mathbb{A}_R \in \{0, 1\}^R$, where $\mathbf{A}_{i,r,j} = 1$ iff the relation (i, r, j) exists in the knowledge graph $\mathbf{A}$. For homogeneous graph without node features, $\mathbb{A}_1 = \{0, 1\}$. W.l.o.g. we define $\mathcal{V}^{(N)} = \{1, \dots, N\}$ and $\mathcal{R}^{(R)} = \{1, \dots, R\}$. If the KG has node attributes, $\mathbb{A}_R$ also encodes them, to be used by the set of self-loops $\{\mathbf{A}_{i,r,i} : i \in \mathcal{V}^{(N)}, r \in \mathcal{R}^{(R,\text{self})}\}$ for a special subset of relations $\mathcal{R}^{(R,\text{self})} \subsetneq \mathcal{R}^{(R)}$. Often $\mathcal{R}^{(R)}$ is described through a bijection to a set of sentences (e.g., $1 \to$ Father, $2 \to$ Grand, $\dots$). *What distinguishes our KG definition from an attributed multigraph (Section 2) is the assumption that the distribution of μ is such that $\mu(\mathbf{A}_G) = \mu(\mathbf{A}_H)$ for any isomorphic KGs $\mathbf{A}_G$ and $\mathbf{A}_H$ ($\mathbf{A}_G \simeq_{KG} H$ as in Definition 4.2).* In this paper we denote this property of μ as *double exchangeability*.

We then define the concept of KG isomorphism as:

Definition 4.2 (KG Isomorphism)**.** We say two multigraphs $\mathbf{A}_G, \mathbf{A}_H \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$ are isomorphic (denoted as $\mathbf{A}_G \simeq_{\text{KG}} \mathbf{A}_H$) if there exists a node bijection $\phi : \mathcal{V}_G \to \mathcal{V}_H$ and a relation bijection $\tau : \mathcal{R}_G \to \mathcal{R}_H$ preserving the set of relations, i.e., $\forall (i, r, j) \in \mathcal{V}_G \times \mathcal{R}_G \times \mathcal{V}_G$, $(\mathbf{A}_G)_{i,r,j} = (\mathbf{A}_H)_{\phi(i),\tau(r),\phi(j)}$.

Remark (vertex and relation set sizes): Note that by Definition 4.1, the set of all knowledge graphs is $\cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$. Common GNN representations can be learned and applied to graphs of different sizes. Similarly, our representations can be learned and applied to KGs with any number of nodes ($N \geq 2$) **and** any number of relations ($R \geq 1$).

Invariant KG representations. It follows from Definition 4.1 that any (statistical) loss function (e.g., likelihood, regression (via energy-based models using distances), cross-entropy) defined over a knowledge graph $\mathbf{A}_G^{\text{train}}$ must be the same over any isomorphic KGs $\mathbf{A}_H \simeq_{\text{KG}} \mathbf{A}_G^{\text{train}}$, i.e., the loss over $\mathbf{A}_G^{\text{train}}$ must be invariant to permutations of the node ids, edge attributes (relations), and node attributes (types). Consequently, we will design representations that are invariant to these two permutations, as we see later.

Definition 4.3 (Permutation actions on KGs)**.** For any KG $\mathbf{A}^{(N,R)} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$. As before, let $\phi \in \mathbb{S}_N$ be an element of the symmetric group $\mathbb{S}_N$ (a permutation). The operation $\phi \circ \mathbf{A}^{(N,R)}$ is the action of ϕ on $\mathbf{A}^{(N,R)}$, defined as

[2] We use $\mathbf{A}$ to denote arbitrary KG of any size instead of $\mathbf{A}^{(N,R)}$, where N and R can be automatically inferred from $\mathbf{A}$.

$(\phi \circ \mathbf{A}^{(N,R)})_{\phi\circ i,r,\phi\circ j} = (\mathbf{A}^{(N,R)})_{i,r,j}, \forall (i,r,j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$. In our definition of KGs, we also need relation permutations, where $\tau \in \mathbb{S}_R$ is a relation permutation and the action of τ on $\mathbf{A}^{(N,R)}$ is defined as $\forall (i,r,j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$, $(\tau \circ \mathbf{A}^{(N,R)})_{i,\tau\circ r,j} = (\mathbf{A}^{(N,R)})_{i,r,j}$, where we define the action of a permutation $\pi \in \mathbb{S}_R$, as $\pi \circ r = \pi_r$.

The node and relation permutation actions on $\mathbf{A}$ are commutative, i.e., $\phi \circ \tau \circ \mathbf{A} = \tau \circ \phi \circ \mathbf{A}$. We now define isomorphic triplets based on the notion of KG isomorphism.

Definition 4.4 (Isomorphic triplets in KGs)**.** $\forall \mathbf{A}_G, \mathbf{A}_H \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$, we say two triplets $(i,r,j) \in \mathcal{V}_G \times \mathcal{R}_G \times \mathcal{V}_G$, $(i',r',j') \in \mathcal{V}_H \times \mathcal{R}_H \times \mathcal{V}_H$ are isomorphic triplets iff $\mathbf{A}_G$ and $\mathbf{A}_H$ have the same graph sizes and relation sizes, and $\exists \phi \in \mathbb{S}_N, \exists \tau \in \mathbb{S}_R$, such that $\phi \circ \tau \circ \mathbf{A}_G = \mathbf{A}_H$.

Definition 4.4 implies isomorphic triplets exists between two isomorphic KGs $\mathbf{A}_G$ and $\mathbf{A}_H$, or in the same graph if $\exists \phi \in \mathbb{S}_N, \exists \tau \in \mathbb{S}_R$ such that $\phi \circ \tau \circ \mathbf{A}_G = \mathbf{A}_G$, where ϕ and τ are not identity maps. Now we can finally define our invariant triplet representation for KGs, which is invariant over isomorphic triplets.

Definition 4.5 (Invariant triplet representation for KGs)**.** For any KG $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$. An invariant representation of a triplet (i,r,j), denoted as $\Gamma_{\text{tri}}((i,r,j),\mathbf{A})$, where $\Gamma_{\text{tri}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} (\mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)} \times \mathbb{A}_R^{N^2}) \to \mathbb{R}^d, d \geq 1$, is such that $\forall (i,r,j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(N)} \times \mathcal{V}^{(N)}, \forall \phi \in \mathbb{S}_N, \forall \tau \in \mathbb{S}_R$, $\Gamma_{\text{tri}}((i,r,j),\mathbf{A}) = \Gamma_{\text{tri}}((\phi\circ i, \tau \circ r, \phi \circ j), \phi \circ \tau \circ \mathbf{A})$.

Remark (quotient group for preserving relations (and node attributes) that do not permute): One can also trivially extend our definitions to restrict exchangeability to a subset of relations. This is achieved by redefining the permutation group $\mathbb{S}_R$ as its quotient group encompassing just the relations that permute, which then implies a trivial change to the definition of KG isomorphism in Definition 4.2. This is a straightforward modification of our approach.

Remark (scoring losses): For $d = 1$, $\Gamma_{\text{tri}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} (\mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)} \times \mathbb{A}_R^{N^2}) \to \mathbb{R}$ can be seen as a *scoring function*, which returns the log-likelihood of an energy-based probability that the corresponding triplet appears in the knowledge graph. For knowledge graph completion tasks, we aim to assign high scores for triplet edges that appears or are missing in the KG. Different from traditional *scoring function* in the literature (Yang et al., 2015; Trouillon et al., 2016; Chen et al., 2022), the invariant triplet representation has additional invariance properties.

Similar to (Srinivasan & Ribeiro, 2020), we can define the most expressive invariant triplet representation.

Definition 4.6 (Most-expressive invariant triplet representation)**.** An invariant triplet representation Γ_{tri} is most expressive iff $\forall \mathbf{A}_G, \mathbf{A}_H \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}, \forall (i,r,j) \in \mathcal{V}_G \times \mathcal{R}_G \times \mathcal{V}_G, \forall (i',r',j') \in \mathcal{V}_H \times \mathcal{R}_H \times \mathcal{V}_H$, we will have $\Gamma_{\text{tri}}((i,r,j),\mathbf{A}_G) = \Gamma_{\text{tri}}((i',r',j'),\mathbf{A}_H)$ iff (i,r,j) and (i',r',j') are isomorphic triplets (Definition 4.4).

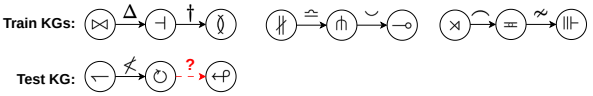


Figure 2. (Alien KG) Illustrative inductive knowledge graph completion task of our alien KG. The task is to inductively predict the missing relation "?" in red. Note that relations are all unique.

In what follows we define representations for the whole KG (akin to how GNNs provide representations for a whole graph), which we denote as *double equivariant KG representations*.

Definition 4.7 (Double equivariant KG representations)**.** Let $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$ be a KG following Definition 4.1. A function $\Gamma_{\text{gra}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2} \to \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{R}^{N\times R\times N\times d}, d \geq 1$ is double equivariant w.r.t. arbitrary node $\phi \in \mathbb{S}_N$ and relation $\tau \in \mathbb{S}_R$ permutations, if $\Gamma_{\text{gra}}(\phi\circ\tau\circ\mathbf{A}) = \phi\circ\tau\circ\Gamma_{\text{gra}}(\mathbf{A})$. Moreover, valid mappings of Γ_{gra} must map a domain element to an image element with the same number of nodes N and relations R.

Next, we connect Definitions 4.5 and 4.7.

Theorem 4.8. *For all $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$, given an invariant triplet representation Γ_{tri} we can construct a double equivariant representation as $(\Gamma_{gra}(\mathbf{A}))_{i,r,j,:} := \Gamma_{tri}((i,r,j),\mathbf{A})$, $\forall (i,r,j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$, and vice-versa.*

Section 5 will introduce a double equivariant neural architecture based on Theorem 4.8. However, first we want to discuss the consequences of invariant representations, and how it can benefit KG tasks.

4.1. Consequences of invariant predictors in KGs

We will now analyze two KG completion tasks that are effectively impossible for all standard KG completion methods (based on attributed multigraphs), which are relatively easy for predictors based on our invariant KG representations.

Consider the knowledge base in Figure 2, obtained from a fictional alien civilization with 3 KGs for training and one for test. Knowing nothing about alien language and costumes, we note that in training all KG relations are different. Minimally, we could predict the missing relation in red in test data is not "$\not<$". Note, however, that because all edge attributes are unique, assuming the KG is an attributed (multi)graph does not allow us to automatically infer this obvious logical rule, since whatever rules are learned for one relation are not directly applicable to others.

Now let's consider sending our aliens a KG with a the biological human relationships in Figure 1(a). Given a set

of biological male relations as training data in Figure 1(a), under what assumptions could the alien be able to predict (without knowledge of our language or physiology) the relation "(Ellie, Grand ∧ Mother, Hanna)" between Ellie and Hanna in the (hold-out) test data of Figure 1(b)?

Thankfully, the tasks in Figures 1 and 2 can both be solved under our definition of KG (Definition 4.1). Due to required invariance, any triplet representation $(i, r, j) \in \mathcal{V} \times \mathcal{R} \times \mathcal{V}$ in either train or test (Definition 4.5) can only pay attention to the structural relations between nodes and their relations, not their absolute ids (node id and relation id). In the KG of Figure 2, any representation invariant to both permutations in training can only encode that any relation is unlike any other relation, that is, a self-supervised trained predictor created by removing the tripled $(\Cap, \smile, \multimap)$ and trying to predict it back must predict a uniform distribution over the remaining relations $\mathcal{R}^{\text{train}} \backslash \{\doteq\}$. If all train KGs are treated as a single (disconnected) KG, the uniform prediction is over $\mathcal{R}^{\text{train}} \backslash \{\Delta, \dagger, \doteq, \frown, \nsim\}$. In test, this predictor would predict the relation "?" uniformly over the set $\mathcal{R}^{\text{test}} \backslash \{\not\asymp\}$, which is really all we know about the aliens.

In the task of Figure 1(a), once we remove (Bob, Grand ∧ Father, Hans) for training (via self-supervision), any invariant triplet predictor for the pair (Bob, Hans) that can correctly predict back the triplet (Bob, Grand ∧ Father, Hans) based on (1-hop) neighbor information from Bob and Hans in training must also be able to predict (Ellie, Grand ∧ Mother, Hanna) in the test KG of Figure 1(b). This is because, restricted to their respective 1-hop neighborhoods, the triplet (Bob, Grand ∧ Father, Hans) in the training KG of Figure 1(a) is isomorphic (Definition 4.4) to the triplet (Ellie, Grand ∧ Mother, Hanna) in the test KG of Figure 1(b).

4.2. Connection to Learning Logical Rules

We now define universally quantified entity and relation Horn clauses for our tasks, and show that any predictor that can be learned from the invariant triplet representation in Definition 4.2 has an equivalent predictor as a conjunction of such Horn clauses.

Definition 4.9 (UQER Horn clauses: Universally quantified entity and relation Horn clauses)**.** We define a subset of universally quantified Horn clauses involving K relations of M entities, defined by an indicator tensor $\mathbf{B} \in \{0,1\}^{M \times K \times M}$:

$$\begin{aligned} &\forall C_1 \in \mathcal{R}, (\forall C_r \in \mathcal{R} \setminus \{C_1, \dots, C_{r-1}\})_{r=2}^{K}, \\ &\forall E_1 \in \mathcal{V}, (\forall E_i \in \mathcal{V} \setminus \{E_1, \dots, E_{i-1}\})_{i=2}^{M}, \\ &\bigwedge_{\substack{i,j=1,\dots,M,\\ r=1,\dots,K,\\ \mathbf{B}_{i,r,j}=1}} (E_i, C_r, E_j) \implies (E_1, C_1, E_h), \end{aligned} \tag{1}$$

for any relation set $\mathcal{R}$ and entity set $\mathcal{V}$ s.t., $|\mathcal{R}| \geq K, |\mathcal{V}| \geq M$, $h \in \{1, 2\}$ (where $h = 1$ indicates a self-loop relation and/or a node attribute).

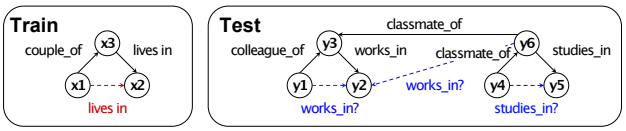


Figure 3. Example of an inductive KG completion task with new relations that can be explained by our Horn clauses.

Note that our definition of UQER Horn clauses (Definition 4.9) is a generalization of the first order logic (FOL) clauses in (Yang et al., 2017; Meilicke et al., 2018; Sadeghian et al., 2019; Teru et al., 2020) such that the relations in the Horn clauses are also universally quantified rather than predefined constants. Note that our Horn clauses need not to form a path in the KG, since some relevant associations between relations could be in disconnected subgraphs.

Figure 3 exemplifies the connection between Definition 4.9 and our KG definition (Definition 4.1). In the training KG, we can see that $(\text{x1}, \text{couple\_of}, \text{x2}) \wedge (\text{x2}, \text{lives\_in}, \text{x3}) \implies (\text{x1}, \text{lives\_in}, \text{x3})$. According to that, we may simply learn that, in a KG, for any two different relations in $\mathcal{R}$ and any three different entities in $\mathcal{V}$, *if they form a logic chain of length* 2 *with distinct relations, then the second relation on the chain also exists between the source and destination entities of the chain*. Using Equation (1) we would write this as $\forall C_1 \in \mathcal{R}, \forall C_2 \in \mathcal{R} \setminus \{C_1\}, \forall E_1 \in \mathcal{V}, \forall E_2 \in \mathcal{V} \setminus \{E_1\}, \forall E_3 \in \mathcal{V} \setminus \{E_1, E_2\}, (E_1, C_1, E_3) \wedge (E_3, C_2, E_2) \implies (E_1, C_2, E_2)$.

Then, on test KG in Figure 3, we will apply the above UQER Horn clause learned from training to predict all missing positive triplets. For instance, an arbitrary variable allocation, "classmate_of", "studies_in" and entities y4, y5, y6, allows all conjunctive conditions of our Horn clause to be satisfied, thus predicting $(\text{y4}, \text{studies\_in}, \text{y5})$ as a positive triplet. Two other triplets can similarly be predicted in dashed blue in Figure 3.

We now connect our double-invariant triplet representations in Definition 4.5 with the UQER Horn clauses in Definition 4.9.

Theorem 4.10. *Given an arbitrary triplet predictor* $\eta : \mathbb{R}^d \to \{0,1\}$ *that takes the triplet representation* $\Gamma_{tri}((i,r,j), \mathbf{A})$ *in Definition 4.5 as input,* $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$*, and predicts if* $(i,r,j) \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$ *is a positive triplet, there exists a set of UQER Horn clauses (Definition 4.9) that predicts the same positive triplets for all* $\mathbf{A} \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2}$ *and* $(i,r,j) \in \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$.

The full proof in Appendix A shows how the universal quantification in Definition 4.9 implies a double-invariant

predictor, where we can construct a set of Horn clauses and for each Horn clause, the left side are observed facts in KG and the right side is triplet predicted to be positive. By adding the UQER property, as a permutation equivariance of nodes and relations, the set of Horn clauses still hold, where the predictor η based on our invariant triplet representations Γ_{tri} (Definition 4.5) also gives the same predictions.

5. Inductive Double-Exchangeable Neural Architecture for KGs

In this section we propose a model to learn invariant triplet representation for KGs. By Theorem 4.8, one way to obtain invariant triplet representation is to learn a double equivariant function (Definition 4.7). So we propose an inductive structural doubly-exchangeable architecture to learn double equivariant functions over KG.

We start by looking at Definition 4.7 from another point of view. Consider $\mathbf{A}^{(N,R)}$ given by Definition 4.1. Denote $A^{(r)}$ as the matrix $A^{(r)}_{i,j} = (\mathbf{A}^{(N,R)})_{i,r,j}$, $r \in \mathcal{R}^{(R)}$. Note that the KG can be written as $\mathbf{A}^{(N,R)} = (A^{(1)}, \dots, A^{(R)})$. Since the actions of the two permutation groups $\mathbb{S}_N$ and $\mathbb{S}_R$ commute, the double equivariance in Definition 4.7 over $\mathbf{A}^{(N,R)}$ can be described as a $\phi \in \mathbb{S}_N$ (graph) equivariance over $A^{(r)}, r = 1, \dots, R$, and a $\tau \in \mathbb{S}_R$ (set) equivariance (over the set of homogeneous graphs). Hence, our double equivariance can make use of the general framework proposed by Maron et al. (2020); Bevilacqua et al. (2021).

We start with a linear double-equivariant layer composed by a Siamese layer to define the k-th linear double-equivariant layer $L^{(t)} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2} \to \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{R}^{N \times R \times N \times d_t}$ as follows, for each $r = 1, ..., R$:

$$(L^{(t)}(\mathbf{A}^{(N,R)}))_{:,r} = L_1^{(t)}(A^{(r)}) + L_2^{(t)}\Big(\sum_{r' \in \mathcal{R} \setminus \{r\}} A^{(r')}\Big), \tag{2}$$

where $t = 1, \dots, T$, $T \geq 2$, $L_1^{(t)}, L_2^{(t)} : \cup_{N=2}^{\infty} \mathbb{A}_1^{N^2} \to \cup_{N=2}^{\infty} \mathbb{R}^{N \times N \times d_t}$ can be any GNN layers that outputs pairwise representations. The sum $\sum_{r' \in \mathcal{R} \setminus \{r\}} A^{(r')}$ can also be replaced by other set aggregators such as mean, max, etc.. Our implementation uses the max aggregator, where $\max\left(\{A^{(r')}\}_{r' \in \mathcal{R} \setminus \{r\}}\right)$ only cares if a pair of nodes is connected (no matter the edge attribute). Note that the proposed layer is similar to the H-equivariant layers proposed by Bevilacqua et al. (2021) for increasing the expressiveness of GNN using sets of subgraphs (a markedly different task than ours). We now can define our (double-equivariant) neural network for KGs:

Definition 5.1 (Double-equivariant neural network)**.** The double-equivariant network $\Gamma_{\text{gra}} : \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{A}_R^{N^2} \to \cup_{R=1}^{\infty} \cup_{N=2}^{\infty} \mathbb{R}^{N \times R \times N \times d}$ is defined by several linear double equivariant layers described in Equation (2) interleaved with non-polynomial activation functions,

$$\Gamma_{\text{gra}}(\mathbf{A}) = L^{(T)}(\cdots \sigma(L^{(2)}(\sigma(L^{(1)}(\mathbf{A}))))\cdots), \tag{3}$$

where σ is the non-polynomial activation function (our implementations uses ReLU).

5.1. Implementation considerations

Most-expressive pairwise representations for $L_1^{(k)}, L_2^{(k)}$ are computationally expensive. Moreover, even less expressive pairwise GNN layers in Equation (2), such as Zhang & Chen (2018); Zhu et al. (2021); Zhang et al. (2021); Zhou et al. (2022), are still expensive (computationally and memory-wise). Thus, we propose inductive structural doubly-exchangeable architecture (IS-DEA), and implementation of Equation (3) that trade-offs expressivity for speed by using equivariant GNN layers (Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018) for node representation. Specifically, for a KG $\mathbf{A}^{(N,R)}$, IS-DEA performs vertex message passing through two learnable functions, such as MLPs, recursively over T layers $\{L^{(t)}\}_{t=1}^{T}$.

At each iteration $t \in \{1, 2, ..., T\}$, all vertices $i \in \mathcal{V}^{(N)}$ are associated with a learned vector $h_i^{(t)} \in \mathbb{R}^{R \times d_t}$, $d_t \geq 1$. Since we do not assume our KGs have node attributes, we consider initializing $h_i^{(0)} = \mathbb{1}$. Then we recursively compute the update, $\forall i \in \mathcal{V}^{(N)}, \forall r \in \mathcal{R}^{(R)}$,

$$\begin{aligned} h_{i,r}^{(t+1)} = \text{MLP}_1^{(t)}\Big(h_{i,r}^{(t)}, \sum_{j \in \mathcal{N}_r(i)} h_{j,r}^{(t)}\Big) \\ + \text{MLP}_2^{(t)}\Big(\sum_{r' \neq r}^{R} h_{i,r'}^{(t)}, \sum_{j \in \cup_{r' \neq r} \mathcal{N}_{r'}(i)} \Big(\sum_{r' \neq r}^{R} h_{j,r'}^{(t)}\Big)\Big), \end{aligned} \tag{4}$$

where $\text{MLP}_1^{(t)}$ and $\text{MLP}_2^{(t)}$ denotes two multi-layer perceptron for the Siamese and aggregation function, $\mathcal{N}_r(i)$ denotes the neighborhood set of i with relation r in the graph, $\mathcal{N}_r(i) = \{j | (i, r, j) \in \mathcal{S} \text{ or } (j, r, i) \in \mathcal{S}\}$ (with $\mathcal{S}$ as the KG triplets encoded by $\mathbf{A}^{(N,R)}$) and $\cup_{j \neq r} \mathcal{N}_r(i)$ denotes the neighborhood set of i in the graph defined by $A^{(r)}$. In our implementation, we use GIN (Xu et al., 2019a) as our GNN architecture, which satisfies Equation (4). At the final layers, we use standard MLPs (which does not take neighborhood information as input) to output a final prediction.

As shown by Srinivasan & Ribeiro (2020); You et al. (2019), structural node representations are not most-expressive for link prediction task in homogeneous graphs. The same issue happens for KGs. To ameliorate the issue, we concatenate i and j (double-equivariant) node representations with the distance between i and j in our triplet representation (appending distances is also adopted in the representations of Teru et al. (2020); Galkin et al. (2021)). Finally, we obtain the triplet representation $\forall (i, r, j) \in \mathcal{V}^{(N)} \times \mathcal{R}^{(R)} \times \mathcal{V}^{(N)}$

$$\Gamma_{\text{IS-DEA}}((i, r, j), \mathbf{A}) = (h_{i,r}^{(T)} \,\|\, h_{j,r}^{(T)} \,\|\, d(i, j) \,\|\, d(j, i)), \tag{5}$$

Model	MRR↑	Hits@1↑	Hits@2↑	Hits@4↑
Neural LP	0.502	0.339	0.415	0.651
DRUM	0.502	0.339	0.415	0.651
GraIL	0.422	0.181	0.416	0.740
NBFNet	0.159	0.168	0.360	0.595
IS-DEA	**0.832**	**0.700**	**0.903**	**1.000**

Table 1. Inductive performance on Family Diagram 1. Existing baselines clearly struggle to perform this task.

Model	MRR↑	Hits@1↑	Hits@2↑	Hits@4↑
Neural LP	N/A	N/A	N/A	N/A
DRUM	N/A	N/A	N/A	N/A
GraIL	N/A	N/A	N/A	N/A
NBFNet	N/A	N/A	N/A	N/A
IS-DEA	**0.915**	**0.839**	**0.974**	**1.000**

Table 2. Inductive Performance on Family Diagram 2. Only our method (IS-DEA) is able to perform this task.

where we denote $d(i, j)$ as the length of the non-trivial shortest path from i to j without direct connection between i and j in the KG, $||$ as the concatenation operation. Since the KG is directed, we concatenate distance on both directions.

Lemma 5.2. *The triplet representation in Equation (5) is an invariant triplet representation as per Definition 4.5.*

Finally, as in previous KG works (Yang et al., 2015; Schlichtkrull et al., 2018; Zhu et al., 2021), we use negative sampling in our training procedure, where for each training triplet $(i, r, j) \in \mathcal{S}$, we randomly corrupt either the subject or object n times to generate the negative example. Following Schlichtkrull et al. (2018), we use cross-entropy loss for model optimization to obtain predictions that will score positive examples higher than negative examples:

$$\mathcal{L} = -\sum_{(i,r,j)\in\mathcal{S}} \Big(\log(\Gamma_{\text{tri}}((i, r, j), \mathbf{A})) - \frac{1}{n}\sum_{p=1}^{n} \log(1 - \Gamma_{\text{tri}}((i'_p, r, j'_p), \mathbf{A})) \Big), \tag{6}$$

where (i'_p, r, j'_p) are the p-th negative examples.

6. Experiments

We evaluate IS-DEA on two synthetic tasks (that we propose to test the generalization capabilities of our method), FD-1 and FD-2; and on two inductive knowledge graph completion datasets, WN18RR-v1 and NELL995-v1, which are widely-used small-scale inductive knowledge graph completion benchmarks in literature (Teru et al., 2020; Zhu et al., 2021). The use of small-scale KGs follows from a limitation of our approach similar to the limitation of GraIL (Teru et al., 2020), where both need to pre-process the graph in order to create egonets for each triplet in the minibatch, a computationally expensive procedure. In all results, we report mean performance over 5 runs, and all variances are omitted since they are quite small which is consistent with the reporting of Teru et al. (2020); Zhu et al. (2021). More experiment details including baselines, implementation details and ablation studies can be found in Appendix D.

6.1. Synthetic Experiments

In the synthetic experiments, we propose two challenging family tree completion tasks in order to verify the theoretical benefits of our model: FD-1 is used to show that our model is insensitive to relation identity; FD-2 is used to show that our model can automatically generalize to new nodes and relations. These tasks are described next.

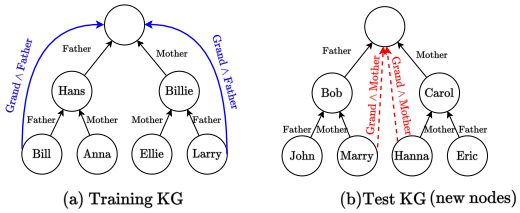


Figure 4. **Simplified Example of FD-1 Generation.** Family tree where negative (positive) training triplets "Grand ∧ Mother" ("Grand ∧ Father") become positive (negative) triplets in test (in this task the KG is finitely exchangeable and a perfect predictor to predict dashed red test edges does not exist), while our assumption is for FD-2 is similar but adds 2 extra relations in the test KG.

Family Diagram 1 (FD-1). A simplified version of FD-1 is illustrated in Figure 4. Given fact and missing positive triplets in a training KG, the goal is to learn a model which can detect missing positive triplets given fact triplets on a different test KG whose relations have different meanings (Appendix D.1 gives more details). The results on FD-1 are shown at Table 1. IS-DEA significantly outperforms all baselines. This task tests the relation-invariance property of IS-DEA.

Interestingly, the baselines that tend to perform better on real-world KGs (e.g., NBFNet (Zhu et al., 2021), GraIL (Teru et al., 2020)) tend to perform worse on FD-1. This is because training and test queries are conflicting: Positive triplet queries in the training graph are negative queries in test, while positive test queries become negative in training. Hence, for models that assume attributed multigraphs (i.e., exchangeable but not double exchangeable multigraphs), the better it can perform on triplets similar to the ones seeing in training, the worse it will perform on the test data.

Family Diagram 2 (FD-2). The FD-2 task is an extension of the scenario described in Figure 3. The learning goal is similar to FD-1. Besides the relation meanings are different from training to test, test KG in FD-2 has more number of relations than training KG. Please refer to Appendix D.1 for further details. On FD-2, training has 127 nodes and 2 relations, while test has 254 nodes and 4 relations (more nodes

Dataset	Model	MRR↑		Hits@1↑		Hits@5↑		**Hits@10↑**	
		Original	Permuted	Original	Permuted	Original	Permuted	Original	Permuted
WN18RR-v1	Neural LP	0.388	0.271	0.179	0.109	0.671	0.437	0.825	0.686
	DRUM	0.395	0.330	0.201	0.157	0.645	0.527	0.823	0.745
	GraIL	0.829	0.789	0.806	0.744	0.840	0.822	0.840	0.822
	NBFNet	0.570	0.273	0.698	0.590	0.817	0.765	0.925	0.906
	IS-DEA	**0.946**	**0.946**	**0.897**	**0.897**	**1.000**	**1.000**	**1.000**	**1.000**
NELL995-v1	Neural LP	0.472	0.455	0.442	0.405	0.500	0.500	0.500	0.500
	DRUM	0.503	0.502	0.500	0.500	0.500	0.500	0.500	0.500
	GraIL	**0.825**	**0.578**	0.737	0.476	0.931	0.706	0.932	0.734
	NBFNet	0.442	0.365	**0.776**	**0.776**	**0.994**	**0.994**	0.998	0.997
	IS-DEA	0.523	0.523	0.506	0.506	0.866	0.866	**1.000**	**1.000**

Table 3. **Results on original and relation-shuffled WN18RR-v1 and NELL995-v1.** IS-DEA outperforms baselines in almost all metrics.

and more relations). Thus, N/A in Table 2 expresses that none of our baselines can perform this task (since, as they assume an attributed multigraph as input, they all need to learn parameters for each relation). Since IS-DEA does not learn parameters specific to relations, it is the only method that can inductively infer over a KG with new and more relations in test, and achieving very good performance on FD-2 as shown in Table 2.

6.2. Real-world Knowledge graphs

As far as we know, there are no real-world benchmarks where training and test KGs have distinct nodes and relations. Therefore, our real-world evaluation of inductive knowledge graph completion is limited to tasks that existing methods can also perform. Unfortunately, due to the complexity of the pre-processing step (similar to GraIL (Teru et al., 2020)) and training cost of the proposed IS-DEA, our experiments are currently limited to small-scale KGs. Thus, we select the smallest two benchmarks, WN18RR-v1 and NELL995-v1 with at most 7,000 fact triplets and 14 relations to test our proposal. In order to highlight the relation-invariance property of our proposal, we also perform a task where all relation IDs are randomly shuffled only in test.

Our results for WN18RR-v1 and NELL995-v1 are reported in Table 3. We can see that IS-DEA results are always invariant to the permutation of relations in test, while all baselines become worse at least on WN18RR-v1 if relations are permuted in test. Besides, IS-DEA obtains a perfect score on the key metric Hits@10, and particularly, is the best of all metrics for WN18RR-v1.

We note in passing that we had to rerun all baselines. For nearly all baselines, we were able to improve their original results on the same benchmarks by better hyperparameter search (except NBFNet in WN18RR-v1). Table 5 reproduces Table 3 with baseline performances taken from original papers. Our conclusions remain the same, as expected.

We also note that WN18RR-v1 and NELL995-v1 are easy tasks that do not test the full capabilities of IS-DEA. Hopefully our work will inspire future benchmark datasets with harder tasks that cannot be performed by existing methods. While double exchangeability may not be the right assumption for all KGs, it is clearly beneficial for some KGs. Our experiments treat all relations as exchangeable. Further research is needed to better understand which relations are exchangeable and which are not for a given KG. We also believe that using true pairwise representation can improve the performance of IS-DEA.

Ablation Study. We also perform an ablation in Table 6 (Appendix) that verifies that, in practice, the shortest path distance features are not essential for IS-DEA in real-world datasets (since real-world KGs are likely asymmetric, where structural node embeddings have similar expressivity to structural pairwise embeddings). IS-DEA is still the best-performing method even without shortest path distances.

Limitations. IS-DEA excels both in synthetic and real-world benchmarks. However, the simplification from pairwise to node embeddings in IS-DEA limits its expressivity. In Appendix D.4, we give a synthetic counterexample how this could be an issue in some KGs. Moreover, IS-DEA has the same poor pre-processing scalability as GraIL. We leave these limitations as future works (see Appendix E). Besides, we do not see any negative social impact of our proposal.

7. Conclusions

In this work we introduced the concept of double exchangeable attributed graphs as a formal model for KGs, challenging the view that KGs are attributed graphs (with exchangeable node ids). We showed that, similar to how node id symmetries impose learning structural node embeddings in homogeneous graphs, double symmetries (node and relation ids) impose structural rule learning in KGs. We then introduced a blueprint for double equivariant neural network architectures for KGs, which adapts permutation-equivariance to both KG entities and relations. We showed this architecture can learn logical rules that standard KG methods cannot. Finally, experiments showed that even a simple double exchangeable architecture (IS-DEA) achieves promising results in inductive KG completion tasks, a significant improvement over baselines.

References

Aldous, D. J. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Barceló, P., Kostylev, E., Monet, M., Pérez, J., Reutter, J., and Silva, J.-P. The logical expressiveness of graph neural networks. In *8th International Conference on Learning Representations (ICLR 2020)*, 2020.

Barcelo, P., Galkin, M., Morris, C., and Orth, M. R. Weisfeiler and leman go relational. *arXiv preprint arXiv:2211.17113*, 2022.

Bevilacqua, B., Frasca, F., Lim, D., Srinivasan, B., Cai, C., Balamurugan, G., Bronstein, M. M., and Maron, H. Equivariant subgraph aggregation networks. In *International Conference on Learning Representations*, 2021.

Bordes, A., Weston, J., Collobert, R., and Bengio, Y. Learning structured embeddings of knowledge bases. In *Twenty-fifth AAAI conference on artificial intelligence*, 2011.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42, 2017.

Chamberlain, B. P., Shirobokov, S., Rossi, E., Frasca, F., Markovich, T., Hammerla, N., Bronstein, M. M., and Hansmire, M. Graph neural networks for link prediction with subgraph sketching. *arXiv preprint arXiv:2209.15486*, 2022.

Chen, Y., Mishra, P., Franceschi, L., Minervini, P., Stenetorp, P., and Riedel, S. Refactor gnns: Revisiting factorisation-based models from a message-passing perspective. In *Advances in Neural Information Processing Systems*, 2022.

Chen, Z., Chen, L., Villar, S., and Bruna, J. Can graph neural networks count substructures? In *Advances in Neural Information Processing Systems*, 2020a.

Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., and Duan, Z. Knowledge graph completion: A review. *Ieee Access*, 8:192435–192456, 2020b.

Cheng, K., Liu, J., Wang, W., and Sun, Y. Rlogic: Recursive logical rule learning from knowledge graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 179–189, 2022.

De Raedt, L. *Logical and relational learning*. Springer Science & Business Media, 2008.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Galárraga, L. A., Teflioudi, C., Hose, K., and Suchanek, F. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 413–422, 2013.

Galkin, M., Denis, E., Wu, J., and Hamilton, W. L. Nodepiece: Compositional and parameter-efficient representations of large knowledge graphs. In *International Conference on Learning Representations*, 2021.

Galkin, M., Zhu, Z., Ren, H., and Tang, J. Inductive logical query answering in knowledge graphs. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Gori, M., Monfardini, G., and Scarselli, F. A new model for learning in graph domains. In *Proceedings. 2005 IEEE international joint conference on neural networks*, volume 2, pp. 729–734, 2005.

Hamilton, W., Ying, Z., and Leskovec, J. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Heckerman, D., Meek, C., and Koller, D. Probabilistic entity-relationship models, prms, and plate models. *Introduction to statistical relational learning*, 2007:201–238, 2007.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Kejriwal, M., Knoblock, C. A., and Szekely, P. *Knowledge graphs: Fundamentals, techniques, and applications*. MIT Press, 2021.

Kersting, K. and De Raedt, L. Bayesian logic programming: theory and tool. *Statistical Relational Learning*, pp. 291, 2007.

Kipf, T. and Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Koller, D. and Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

Koller, D., Friedman, N., Džeroski, S., Sutton, C., McCallum, A., Pfeffer, A., Abbeel, P., Wong, M.-F., Meek, C., Neville, J., et al. *Introduction to statistical relational learning*. MIT press, 2007.

Kroetsch, M. and Weikum, G. Special issue on knowledge graphs. *Journal of Web Semantics*, 37(38):53–54, 2016.

Lao, N. and Cohen, W. W. Relational retrieval using a combination of path-constrained random walks. *Machine learning*, 81(1):53–67, 2010.

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

Maron, H., Ben-Hamu, H., Serviansky, H., and Lipman, Y. Provably powerful graph networks. In *Advances in Neural Information Processing Systems*, pp. 2156–2167, 2019.

Maron, H., Litany, O., Chechik, G., and Fetaya, E. On learning sets of symmetric elements. In *International Conference on Machine Learning*, pp. 6734–6744. PMLR, 2020.

Meilicke, C., Fink, M., Wang, Y., Ruffinelli, D., Gemulla, R., and Stuckenschmidt, H. Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion. In *International semantic web conference*, pp. 3–20. Springer, 2018.

Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., and Grohe, M. Weisfeiler and leman go neural: Higher-order graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 4602–4609, Jul. 2019.

Murphy, R., Srinivasan, B., Rao, V., and Ribeiro, B. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. In *International Conference on Learning Representations*, 2019a.

Murphy, R., Srinivasan, B., Rao, V., and Ribeiro, B. Relational pooling for graph representations. In *International Conference on Machine Learning*, pp. 4663–4673. PMLR, 2019b.

Murphy, R., Srinivasan, B., Rao, V., and Ribeiro, B. Relational pooling for graph representations. In *Proceedings of the 36th International Conference on Machine Learning*, 2019c.

Neville, J. and Jensen, D. Relational dependency networks. *Journal of Machine Learning Research*, 8(3), 2007.

Nickel, M., Tresp, V., and Kriegel, H.-P. A three-way model for collective learning on multi-relational data. In *Icml*, 2011.

Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2015.

Nickel, M., Rosasco, L., and Poggio, T. Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Paulheim, H. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3): 489–508, 2017.

Qian Huang, H. R. and Leskovec, J. Few-shot relational reasoning via connection subgraph pretraining. In *Neural Information Processing Systems*, 2022.

Raedt, L. D., Kersting, K., Natarajan, S., and Poole, D. Statistical relational artificial intelligence: Logic, probability, and computation. *Synthesis lectures on artificial intelligence and machine learning*, 10(2):1–189, 2016.

Ruffinelli, D., Broscheit, S., and Gemulla, R. You can teach an old dog new tricks! on training knowledge graph embeddings. In *International Conference on Learning Representations*, 2020.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

Sadeghian, A., Armandpour, M., Ding, P., and Wang, D. Z. Drum: End-to-end differentiable rule mining on knowledge graphs. *Advances in Neural Information Processing Systems*, 32, 2019.

Schlichtkrull, M., Kipf, T. N., Bloem, P., Berg, R. v. d., Titov, I., and Welling, M. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.

Schneider, E. W. Course modularization applied: The interface system and its implications for sequence control and data analysis. 1973.

Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., Uribe, J. F. C., Fedus, L., Metz, L., Pokorny, M., Lopes, R. G., Zhao, S., Vijayvergiya, A., Sigler, E.,

Perelman, A., Voss, C., Heaton, M., Parish, J., Cummings, D., Nayak, R., Balcom, V., Schnurr, D., Kaftan, T., Hallacy, C., Turley, N., Deutsch, N., Goel, V., Ward, J., Konstantinidis, A., Zaremba, W., Ouyang, L., Bogdonoff, L., Gross, J., Medina, D., Yoo, S., Lee, T., Lowe, R., Mossing, D., Huizinga, J., Jiang, R., Wainwright, C., Almeida, D., Lin, S., Zhang, M., Xiao, K., Slama, K., Bills, S., Gray, A., Leike, J., Pachocki, J., Tillet, P., Jain, S., Brockman, G., and Ryder, N. ChatGPT: Optimizing language models for dialogue. Official OpenAI Blog, November 2022.

Schuster, M. and Paliwal, K. K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

Shen, T., Zhang, F., and Cheng, J. A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems*, pp. 109597, 2022.

Singhal, A. Introducing the Knowledge Graph: things, not strings. Official Google Blog, May 2012.

Srinivasan, B. and Ribeiro, B. On the equivalence between positional node embeddings and structural graph representations. In *Eighth International Conference on Learning Representations*, 2020.

Sun, Z., Hu, W., Zhang, Q., and Qu, Y. Bootstrapping entity alignment with knowledge graph embedding. In *IJCAI*, volume 18, 2018.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2019.

Sun, Z., Wang, C., Hu, W., Chen, M., Dai, J., Zhang, W., and Qu, Y. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 222–229, 2020.

Sutskever, I., Tenenbaum, J., and Salakhutdinov, R. R. Modelling relational data using bayesian clustered tensor factorization. *Advances in neural information processing systems*, 22, 2009.

Teru, K., Denis, E., and Hamilton, W. Inductive relation prediction by subgraph reasoning. In *International Conference on Machine Learning*, pp. 9448–9457. PMLR, 2020.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. Complex embeddings for simple link prediction. In *International conference on machine learning*, pp. 2071–2080. PMLR, 2016.

Trouillon, T., Dance, C., Gaussier, E., Welbl, J., Riedel, S., and Bouchard, G. Knowledge graph completion via complex tensor factorization. *Journal of Machine Learning Research*, 18(130):1–38, 2017.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *ICLR*, 2018.

Wang, H., Ren, H., and Leskovec, J. Relational message passing for knowledge graph completion. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 1697–1707, 2021.

Wang, Z., Zhang, J., Feng, J., and Chen, Z. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014.

Wang, Z., Lv, Q., Lan, X., and Zhang, Y. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 349–357, 2018.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019a.

Xu, K., Wang, L., Yu, M., Feng, Y., Song, Y., Wang, Z., and Yu, D. Cross-lingual knowledge graph alignment via graph matching neural network. In *Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics (ACL), 2019b.

Yan, Y., Liu, L., Ban, Y., Jing, B., and Tong, H. Dynamic knowledge graph alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4564–4572, 2021.

Yang, B., Yih, S. W.-t., He, X., Gao, J., and Deng, L. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*, 2015.

Yang, F., Yang, Z., and Cohen, W. W. Differentiable learning of logical rules for knowledge base reasoning. *Advances in neural information processing systems*, 30, 2017.

You, J., Ying, R., and Leskovec, J. Position-aware graph neural networks. In *International conference on machine learning*, pp. 7134–7143. PMLR, 2019.

Zhang, M. and Chen, Y. Link prediction based on graph neural networks. *Advances in neural information processing systems*, 31, 2018.

Zhang, M., Li, P., Xia, Y., Wang, K., and Jin, L. Labeling trick: A theory of using graph neural networks for multi-node representation learning. *Advances in Neural Information Processing Systems*, 34:9061–9073, 2021.

Zhou, Y., Kutyniok, G., and Ribeiro, B. OOD link prediction generalization capabilities of message-passing GNNs in larger test graphs. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022.

Zhu, Z., Zhang, Z., Xhonneux, L.-P., and Tang, J. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in Neural Information Processing Systems*, 34:29476–29490, 2021.