Research

A Novel Graph-Based Approach Using Graph Neural Networks for Structuring Large-Scale Commonsense Knowledge Bases

Wei Zhang[1], **Lian Chen**[2]

[1] Xinyang Agricultural and Forestry University, Department of Computer Science, Xinyang, Henan, China
[2] Qiannan Normal University for Nationalities, School of Information Engineering, Duyun, Guizhou, China

Abstract: Commonsense knowledge bases (CSKBs) represent real-world facts through entities and relationships, yet their often unstructured or semi-structured nature impedes efficient scalability and reasoning. This paper proposes a novel graph neural network (GNN)-based framework designed to autonomously organize large-scale CSKBs into semantically coherent, machine-interpretable graphs. The core architecture incorporates heterogeneous graph convolution operators to simultaneously capture multi-relational dependencies and preserve local topological features. In tandem, a hierarchical attention mechanism adjusts edge weights dynamically based on node attribute similarity and global graph connectivity, leading to efficient discovery of sparse substructures. To mitigate the combinatorial explosion of relational paths, we introduce a constrained optimization objective that minimizes edge reconstruction loss while maximizing deductive closure through implicit Horn clause satisfiability. Empirical evaluations on ConceptNet and ATOMIC confirm notable improvements in multiple metrics: a 23.7% boost in edge prediction accuracy over competitive graph autoencoder baselines and a semantic consistency score of 0.892 on held-out triples, outperforming transformer-based knowledge base completion methods by 15.2%. Qualitative structural analyses reveal emergent hypernymy and causality hierarchies without explicit ontological supervision, highlighting the robustness of the learned graph representations. Our results emphasize that unifying geometric embeddings with symbolic reasoning constraints significantly enhances the structuring of noisy commonsense assertions, thus promoting scalable, high-fidelity CSKBs.

1. Introduction

Commonsense knowledge bases (CSKBs) serve as repositories of everyday facts and inferences essential for numerous artificial intelligence (AI) tasks, spanning natural language understanding, robotic perception, and automated reasoning. These repositories typically encompass millions of loosely structured assertions, each encapsulating a plausible real-world relation, such as "dog hasA tail" or "rain causes wet ground." The construction of CSKBs often integrates multiple methodologies, including crowdsourcing, automated text mining, and expert annotation, leading to a mixture of redundancies, noise, and partial inconsistencies [1][2] [3]. A significant challenge in maintaining and utilizing CSKBs arises from their inherent relational polysemy, wherein a single relation type, such as "relatedTo" in ConceptNet, can signify vastly different semantic connections. Additionally, entity occurrences within CSKBs often adhere to long-tailed distributions, wherein a small subset of concepts appears frequently, while a vast majority remain sparsely represented. Traditional structuring techniques, including manually curated taxonomies and embedding-based alignment approaches, encounter persistent obstacles in addressing relational heterogeneity and ensuring deductive completeness, especially as CSKBs expand in scale [4] [5] [6].

Recent advancements in graph-based learning methodologies, particularly Graph Neural Networks (GNNs), provide a promising avenue for leveraging CSKBs' rich relational structures. GNNs operate by iteratively aggregating node information through

. . *Helex-science* **2021**, *6*, 1–16.

message-passing mechanisms, thereby capturing both topological dependencies and node-specific features. However, conventional GNN architectures, such as Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), are primarily designed for relatively homogeneous graphs, where node interactions exhibit a degree of consistency in their semantic interpretations. This assumption does not hold in CSKBs, where relations like "causes" fundamentally differ in nature from relations like "desires," "motivates," or "isA." Knowledge graph-oriented adaptations, such as Relational Graph Convolutional Networks (RGCN), introduce mechanisms to handle multi-relation scenarios by employing relation-specific transformation matrices. Yet, these models often impose rigid schema constraints that restrict their ability to generalize to new or ambiguous relations, limiting their adaptability in evolving CSKB landscapes [7] [8] [9,10].

A key challenge in integrating GNNs with CSKBs lies in the trade-off between structural flexibility and semantic precision. While conventional knowledge graphs (KGs) benefit from well-defined ontological schemas, CSKBs embody a broader and more loosely defined semantic space. This discrepancy necessitates more expressive modeling techniques capable of dynamically adapting to relation-specific variations. One potential approach involves the incorporation of heterogeneous graph neural networks (HGNNs), which explicitly account for diverse node and edge types by introducing type-aware aggregation mechanisms. Additionally, attention-based models, such as Relational Graph Attention Networks (RGAT), can selectively weigh relational influences, thereby mitigating the oversmoothing effects commonly observed in deep GNN architectures. The ability to differentiate between relation-specific influences is particularly crucial in CSKBs, where high-connectivity nodes often serve as semantic hubs that bridge disparate conceptual clusters.

Furthermore, representation learning strategies tailored to CSKBs must address sparsity and long-tailed distributions. Many real-world CSKBs exhibit a power-law distribution in entity occurrences, with a small subset of frequently referenced concepts dominating the knowledge space while a vast majority of entities remain underrepresented. Standard node embedding techniques struggle to provide meaningful representations for such infrequent entities due to limited neighborhood information. One potential remedy involves the incorporation of meta-learning strategies, wherein models are trained to generalize from a few-shot learning paradigm, effectively enhancing their capacity to infer relations for low-resource entities. Additionally, contrastive learning frameworks, which maximize representation distinctions between semantically disparate entities while reinforcing similarities within conceptually related groups, have demonstrated potential in mitigating the adverse effects of data sparsity [11] [12] [13] [14].

Beyond structural and representation challenges, reasoning over CSKBs introduces another layer of complexity. While symbolic reasoning approaches, such as rule-based inference engines, offer high interpretability, they often suffer from scalability constraints and limited generalization to novel assertions. Conversely, embedding-based reasoning frameworks leverage vectorized representations to infer implicit relationships; however, they typically lack transparency in their decision-making processes. Hybrid reasoning models, which integrate symbolic logic with neural embeddings, represent a promising direction for enhancing both interpretability and adaptability. For instance, neuro-symbolic models that incorporate differentiable rule induction mechanisms can dynamically infer logical relationships while preserving explainability. Such models hold particular promise for downstream applications, including commonsense question answering, contextual reasoning in dialogue systems, and autonomous decision-making in robotics [15] [16] [17,18].

The evaluation of CSKB structuring methodologies necessitates robust benchmarking frameworks that account for both structural integrity and semantic coherence. Traditional knowledge graph benchmarks, such as link prediction and triplet classification, provide useful but insufficient measures of performance due to their focus on syntactic rather than conceptual correctness. More comprehensive evaluation paradigms should incorporate multi-faceted assessments, including logical consistency checks, human validation studies, and downstream task performance metrics. Furthermore, introducing adversarial perturbations in benchmark datasets can help assess the robustness of CSKB-enhanced models against erroneous or misleading assertions [19] [20].

Method	Strengths	Weaknesses
Manual Taxonomies	High precision, structured knowledge	Scalability limitations, labor-intensive
Embedding-Based Alignment	Efficient for large-scale integration	Lacks interpretability, struggles with polysemy
Graph Neural Networks (GNNs)	Captures relational dependencies	Sensitive to sparsity, oversmoothing risks
Relational Graph Neural Networks (RGCNs)	Handles multi-relation scenarios	Rigid schema constraints, limited adaptability
Hybrid Neuro-Symbolic Models	Balances interpretability and adaptability	Computationally complex, requires fine-tuning

Table 1. Comparison of Different Approaches for Structuring Commonsense Knowledge Bases

Emerging techniques in self-supervised learning (SSL) further augment the capabilities of GNN-based CSKB models. Self-supervised pretraining strategies, such as masked concept prediction and contrastive graph learning, enable models to develop more generalized representations without requiring extensive labeled data. Additionally, reinforcement learning (RL) paradigms have been explored for optimizing knowledge graph traversal, wherein an RL agent learns optimal paths for information retrieval based on reward mechanisms. These advancements suggest promising directions for enhancing the scalability and effectiveness of CSKB structuring methodologies [21] [22] [23] [24]

Simultaneously, commonsense reasoning frequently involves chaining multiple relational steps: for instance, inferring that "person X wants to eat" might lead to "person X goes to a restaurant," which in turn implies "person X might pay the bill." Such transitive and causal links can explode combinatorially if the structuring algorithm indiscriminately includes all plausible connections. To handle this complexity, deductive logic constraints—represented by Horn clauses or other forms of logical implication—can guide the model to focus on consistent, essential connections [25] [26] [27]. .

In this paper, we propose a unified framework that combines geometric deep learning and symbolic reasoning to address these challenges. On the geometric side, we employ an edge-conditioned GNN that jointly models node embeddings and relation-specific transformations to capture how different relations affect neighbor aggregation. We enhance expressivity and pruning capability through a hierarchical attention mechanism that dynamically modulates edge importance. On the symbolic side, we introduce a deductive regularization term that encourages the discovered edges to satisfy common Horn-clause constraints inherent in many commonsense assertions (e.g., "if living being is dog and dog is an animal, then living being is animal"). By integrating these components, our framework achieves scalable structuring of massive CSKBs, ensuring both representational fidelity (i.e., reconstructing edges accurately) and robust deductive closures (i.e., generating missing but logically implied edges).

Crucially, our model also adapts to hierarchical and associative relations by learning curvature parameters in a hyperbolic or Lorentzian manifold. This accommodates the often tree-like nature of "hypernymy" or "causes" relations, which display hierarchical layering in conceptual space. Through this dynamic, we avoid artificially constraining all relations to a purely Euclidean manifold, where hierarchical distance measures might be suboptimal [28] [29].

To validate the proposed method, we conduct extensive experiments on two large-scale commonsense resources: ConceptNet 5.7 and ATOMIC 2020. Our architecture demonstrates marked improvements over state-of-the-art graph embedding and knowledge graph completion baselines, including TransE, ComplEx, and RGCN variants. Metrics such as edge prediction, semantic consistency of newly inferred edges, and deductive closure rates all favor our approach. Furthermore, an ablation study reveals the critical roles of dynamic edge weighting and symbolic regularization; removing either component significantly reduces performance. A deeper inspection of the learned graphs shows emergent structured hierarchies that resemble manually created taxonomies, despite the complete absence of explicit hierarchical supervision [30,31] [32] [33].

In the sections that follow, we provide a comprehensive examination of the proposed model. We begin by formalizing the structuring task, detailing the notations and constraints we impose on the CSKB. Next, we describe our hybrid neural architecture and its training objective, which unifies geometric embedding, attention-based pruning, and deductive logic constraints. We then present experimental protocols, datasets, and thorough quantitative and qualitative evaluations. Subsequently, we delve into theoretical aspects of model expressiveness, relating it to the Weisfeiler-Lehman test for graph isomorphism and discussing the implications of hyperbolic geometry for hierarchical data. We conclude with a summary of our findings, possible extensions to temporal or modality-rich commonsense knowledge, and considerations for synergy with large-scale language models [34] [35] [36].

2. Methodology

2.1. Problem Formulation and Structured Representation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ denote a labeled multigraph representing the commonsense knowledge base. Each node $v_i \in \mathcal{V}$ corresponds to an entity or concept (e.g., "dog," "rain," "happy"), and each edge $e_{ij}^r \in \mathcal{E}$ represents a relational assertion of type $r \in \mathcal{R}$ between nodes v_i and v_j. These relations may include "isA," "hasA," "causes," "desires," or any other label found in the CSKB. Additionally, nodes often carry feature vectors $\mathbf{x}_i \in \mathbb{R}^d$, either derived from textual descriptions, word embeddings, or specialized concept embeddings. We aggregate these into $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d}$.

A key challenge is that $\mathcal{E}$ can be extremely large, uncurated, and noisy. Edges are often extracted from multiple sources with varying confidence levels. Formally, each edge e_{ij}^r includes a confidence score $w_{ij}^r \in [0, 1]$. The goal of "structuring" the CSKB is to map $\mathcal{G}$ to a refined graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, where $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{E}'$ reflects a pruned and augmented set of edges. This refined graph should maintain high accuracy in reconstructing genuine relations while also supporting deductive inference—i.e., if a fact can logically be derived from existing edges, it should appear (or be inferrable) in $\mathcal{G}'$. More formally:

1. **Completeness**: For any triple (v_i, r, v_j) derivable via a set of Horn clauses on $\mathcal{G}$, there should exist a corresponding edge $e_{ij}^r \in \mathcal{E}'$ (or at least be deducible from a path in $\mathcal{E}'$). 2. **Consistency**: Mutually exclusive relations—such as "isA cat" and "isA dog" if the domain logic forbids intersection—should not both appear on the same node pair without justification. 3. **Conciseness**: Redundant edges that can be inferred transitively (e.g., "X isA mammal" if "X isA dog" and "dog isA mammal" are already present) should ideally be pruned or assigned minimal weight, ensuring the graph remains interpretable and less cluttered.

2.2. Hybrid Graph Neural Network Architecture

Our proposed approach incorporates two parallel streams of message passing: an edge-conditioned convolution pathway and an attention-based mechanism. The overarching intuition is that edges of different types can exhibit substantially different aggregation patterns (e.g., "similarTo" might average neighbor representations, while "isA" might shift embeddings in a more hierarchical fashion), and these variations must be captured explicitly. Meanwhile, we also need a flexible attention mechanism that can dynamically reweight or remove edges that are deemed uninformative or spurious [37] [38] [39].

Edge-Conditioned Convolution Stream

For each node v_i, let $\mathbf{h}_i^{(l)}$ be its embedding at layer l. We define a relation-specific transformation for each $r \in \mathcal{R}$, with learnable parameters $\mathbf{W}_r^{(l)}, \mathbf{b}_r^{(l)}$. The edge-conditioned update is:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_i^r} \frac{1}{|\mathcal{N}_i^r|} \left(\mathbf{W}_r^{(l)}\, \mathbf{h}_j^{(l)} + \mathbf{b}_r^{(l)}\right)\right),$$

where $\mathcal{N}_i^r$ is the set of neighbors connected to v_i by edges of type r. This formulation is reminiscent of relational GCNs but extends them by allowing a more granular weighting structure (discussed further in the next subsection). The activation σ can be ReLU or another nonlinear function.

Attention-Based Convolution Stream

In parallel, we compute an attention-based update akin to GAT:

$$\alpha_{ij} = \mathrm{softmax}_{j\in\mathcal{N}_i}\Big(\mathrm{LeakyReLU}\big(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i^{(l)} \,\|\, \mathbf{W}\mathbf{h}_j^{(l)}]\big)\Big),$$

$$\mathbf{h}_i^{(l+1)} = \sigma\Big(\sum_{j\in\mathcal{N}_i} \alpha_{ij}\, \mathbf{W}\, \mathbf{h}_j^{(l)}\Big),$$

where $\mathcal{N}_i$ can include all neighbors or a relation-specific subset. The key difference here is that α_{ij} is learned adaptively, allowing the model to focus on the most salient neighbors.

Fusion and Gating

We fuse the outputs of these two parallel streams at each layer using a gated mixture parameter β:

$$\mathbf{H}^{(l+1)} = \beta \cdot \mathbf{H}_{\text{edge}}^{(l+1)} + (1-\beta) \cdot \mathbf{H}_{\text{attn}}^{(l+1)},$$

where $\mathbf{H}^{(l+1)}$ is the final embedding matrix at layer $l+1$, $\mathbf{H}_{\text{edge}}^{(l+1)}$ represents the edge-conditioned convolution outputs, and $\mathbf{H}_{\text{attn}}^{(l+1)}$ represents the attention-based outputs. This gating allows the network to balance the specialized relation-based transformations with the more flexible attention mechanisms.

Edge Scoring and Relation-Specific Matrices

After the final GNN layer, we obtain $\mathbf{h}_i$ for each node v_i. We then predict whether an edge e_{ij}^r should exist (or remain) using a relation-specific bilinear scoring function:

$$\phi(e_{ij}^r) = \mathrm{sigmoid}\Big(\mathbf{h}_i^T \mathbf{M}_r\, \mathbf{h}_j + c_r\Big),$$

where $\mathbf{M}_r \in \mathbb{R}^{d\times d}$ and $c_r \in \mathbb{R}$ are learnable parameters. A high $\phi(e_{ij}^r)$ indicates that the relation r between v_i and v_j is likely valid. This score thus drives the reconstruction of $\mathcal{E}$ and simultaneously supports the filtering out of low-confidence edges.

2.3. Dynamic Edge Weighting via Curvature Learning

To handle the inherent hierarchy found in many commonsense relations—such as hypernymy or causality—we allow the model to map nodes into a curved manifold, typically hyperbolic or Lorentzian space, as this geometry often better captures tree-like structures. Let $\mathbf{z}_i \in \mathbb{D}^d$ represent the node embedding in a hyperbolic space of dimension d. The primary motivation behind employing such non-Euclidean spaces is their ability to embed hierarchical structures with lower distortion compared to traditional Euclidean embeddings. Specifically, hyperbolic spaces provide exponentially increasing volumes, allowing them to naturally fit tree-like structures with minimal loss. This property makes them particularly well-suited for knowledge graphs, taxonomies, and other relational data that exhibit strong hierarchical properties. In this setting, we compute an adaptive weight:

$$w_{ij}^{r} = \sigma\Big(\gamma \cdot d_{\mathcal{L}}(\mathbf{z}_i, \mathbf{z}_j) + \delta_r\Big),$$

where $d_{\mathcal{L}}(\cdot, \cdot)$ denotes Lorentzian distance, γ is a learned global curvature parameter, and δ_r is a relation-specific shift. Intuitively, if two nodes are "close" in hyperbolic space, they should be connected more tightly in the final structure. The Lorentzian distance is defined as:

$$d_{\mathcal{L}}(\mathbf{z}_i, \mathbf{z}_j) = \operatorname{arcosh}\big(-\langle \mathbf{z}_i, \mathbf{z}_j \rangle_{\mathcal{L}}\big),$$

where $\langle \mathbf{z}_i, \mathbf{z}_j \rangle_{\mathcal{L}}$ represents the Minkowski inner product in Lorentzian space. This formulation enables hierarchical relationships to be captured naturally, as distances grow logarithmically with depth in the hierarchy.

Relations implying hierarchy (e.g., "isA," "partOf," "causes") often exhibit a characteristic geometry with one side more "central" than the other. That is, the hyperbolic representation places more general concepts near the origin, while more specific instances are positioned toward the periphery. This ordering ensures that child nodes inherit properties from their parents while maintaining appropriate distances for effective generalization. The dynamic weighting introduced via w_{ij}^{r} helps the GNN layers scale properly in highly hierarchical or tree-like subgraphs, while still allowing more associative relations (e.g., "relatedTo") to remain relatively flat. The ability to adaptively learn curvature via γ is crucial, as different knowledge graph domains exhibit varying degrees of hierarchy, and a fixed curvature assumption may not generalize well across tasks.

Incorporation into the GNN

We integrate w_{ij}^{r} by reinterpreting the neighbor aggregation steps. Specifically, the sums over neighbors become:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_i^r} w_{ij}^{r} \Big(\mathbf{W}_r^{(l)}\,\mathbf{h}_j^{(l)} + \mathbf{b}_r^{(l)}\Big)\right),$$

instead of the uniform $\frac{1}{|\mathcal{N}_i^r|}$. This approach more finely controls how strongly each neighbor contributes to the updated embedding, based on both confidence scores and geometric proximity in the learned manifold. The learned weight w_{ij}^{r} introduces anisotropic influences, where nodes closer in hyperbolic space contribute more significantly to the updated representation. This property aligns well with real-world relational graphs, where certain edges (e.g., hypernyms) should be weighted more heavily than others.

Furthermore, the parameterization of δ_r allows relation-specific tuning of influence. For example, in a biological taxonomy, the relation "isA" may have a different optimal shift than a looser semantic relationship such as "relatedTo." The ability to distinguish these cases dynamically makes the approach flexible across different domains.

In practical terms, the use of hyperbolic space in message passing networks has shown improvements in representation quality, especially in scenarios where hierarchical information is crucial. The use of non-Euclidean aggregation aligns naturally with real-world graphs, particularly knowledge bases that exhibit ontological structure. Theoretical guarantees from Riemannian geometry suggest that distances in hyperbolic space maintain meaningful separation, which translates into improved clustering of conceptually similar nodes.

Additionally, the learned curvature parameter γ plays a significant role in determining how aggressively the model distinguishes hierarchical levels. A large negative γ corresponds to extreme hierarchical separation, whereas a smaller magnitude implies a flatter embedding structure. This flexibility allows adaptation across different datasets, ranging

Relation Type	**Curvature Tendency**	**Weighting Impact**	**Example**
Hypernymy ("isA")	Strongly Negative	High Influence	Animal $\rightarrow$ Mammal
Part-Whole ("partOf")	Moderately Negative	Medium Influence	Wheel $\rightarrow$ Car
Causal ("causes")	Variable	Context-Dependent	Fire $\rightarrow$ Smoke
Associative ("relatedTo")	Near Zero	Low Influence	Coffee $\rightarrow$ Caffeine

Table 2. Impact of Relation Type on Curvature and Weighting

from deeply structured taxonomies to relatively shallow knowledge graphs [40,41] [42] [43].

2.4. Multi-Task Learning Objective

We optimize three primary objectives simultaneously to ensure that (1) the proposed structure remains faithful to the original CSKB, (2) logical constraints are not violated, and (3) the resulting subgraph is sufficiently sparse for interpretability.

(1) Reconstruction Loss

We use a contrastive margin-based loss to distinguish observed edges from negative edges:

$$\mathcal{L}_{\text{rec}} = \sum_{(i,j,r)\in\mathcal{E}} \sum_{(i',j',r')\in\mathcal{E}_{\text{neg}}} \max\Big(0, \phi(e_{ij}^r) - \phi(e_{i'j'}^{r'}) + \lambda\Big),$$

where $\mathcal{E}_{\text{neg}}$ is a set of corrupted or sampled negative edges, and λ is the margin hyperparameter. This encourages $\phi(e_{ij}^r)$ for true edges to exceed scores for false edges by a margin.

(2) Deductive Regularization

We introduce a penalty that enforces basic logical constraints via Horn clauses of the form:

$$e_{ik}^s \leftarrow e_{ij}^r \wedge e_{jk}^t,$$

meaning if (v_i, r, v_j) and (v_j, t, v_k) both hold, then (v_i, s, v_k) should likely hold. We translate this to a differentiable constraint:

$$\mathcal{L}_{\text{logic}} = \mathbb{E}_{(e_{ik}^s \leftarrow e_{ij}^r \wedge e_{jk}^t)}\Big[\max\big(0,\ \phi(e_{ik}^s) - \min(\phi(e_{ij}^r), \phi(e_{jk}^t))\big)\Big].$$

If $\phi(e_{ij}^r)$ and $\phi(e_{jk}^t)$ are both large, but $\phi(e_{ik}^s)$ is small, the model incurs a penalty. This pushes it toward consistency.

(3) Topological Sparsity

We add an ℓ_1-penalty on the attention coefficients or dynamic weights to encourage pruning:

$$\mathcal{L}_{\text{sparsity}} = \sum_{r\in\mathcal{R}} \sum_{(i,j)} \|w_{ij}^r\|_1.$$

Weighting this appropriately reduces the connectivity of the final graph, focusing on essential edges.

Unified Objective

We combine these terms as follows:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \alpha\,\mathcal{L}_{\text{logic}} + \beta\,\mathcal{L}_{\text{sparsity}},$$

where α and β control the relative importance of logical consistency and sparsity, respectively. By tuning these hyperparameters, we can emphasize or de-emphasize certain structural properties of the final graph.

2.5. Algorithmic Implementation

Algorithmically, we adopt mini-batch training. Nodes and edges are sampled in subgraphs, ensuring that relevant neighbors for each node are included. We compute the GNN forward pass, derive $\phi(e^r_{ij})$ for both positive and negative edges, and backpropagate through the combined loss. We alternate between generating negative edges $\mathcal{E}_{\text{neg}}$ using random or proximity-based corruption strategies to stabilize training. A specialized sampler picks Horn clauses from the existing subgraph to populate $\mathcal{L}_{\text{logic}}$. Over successive epochs, the GNN parameters, relation-specific transformations, curvature parameters, and gating factors all co-adapt to yield a refined embedding space. This approach is scalable to millions of edges by leveraging efficient GPU-based message passing and negative sampling routines [44] [45] [41,46].

3. Experimental Evaluation

3.1. Datasets

We conduct experiments on two prominent CSKBs:

ConceptNet 5.7

Contains around 1.2 million edges linking over 800,000 concepts with relations like "isA," "hasA," "partOf," "capableOf," "desires," and many more. The graph is multilingual and includes crowd-sourced as well as text-mined edges with varying confidence scores.

ATOMIC 2020

Focuses on if-then reasoning about everyday events, featuring relations such as "xIntent," "xEffect," "xNeed," "xWant," etc. There are roughly 1.3 million edges in total, each representing a causal or motivational aspect of a scenario. ATOMIC edges are typically more abstract and revolve around event-based reasoning.

We filter these datasets to ensure each relation type has a sufficient number of edges and split them into training, validation, and test sets. Negative edges are generated by randomly permuting nodes or sampling from low-confidence pairs that do not appear in the original graph.

3.2. Baselines and Comparisons

We compare against several state-of-the-art methods:

- **TransE**: A classical knowledge embedding method that represents relations as translations in a Euclidean space. - **ComplEx**: Captures relational patterns using complex-valued embeddings, providing better handling of symmetric and antisymmetric relations. - **RGCN**: A relational GNN that applies distinct transformations for different relation types. - **GraphSAGE**: Learns to sample and aggregate neighbor features; we adapt it to handle multi-relational data by simply concatenating relation embeddings. - **KG-BERT / BERT-based**: A transformer-based approach that treats each triple as a textual input for classification, sometimes used for knowledge base completion.

Our method differs by integrating dynamic edge weighting in hyperbolic space, hierarchical attention, and deductive constraints under one framework.

3.3. Evaluation Metrics

We employ three main metrics:

1. **Edge Prediction AUC-ROC**: For each triple in the test set, we assess whether $\phi(e_{ij}^r)$ ranks it above negative edges. 2. **Semantic Consistency**: We randomly sample 500 new edges inferred by the model (i.e., edges with no direct counterpart in the training set) and ask human annotators to judge their plausibility. 3. **Deductive Closure Rate**: We measure the percentage of logically implied triples (via Horn clauses) that the structured graph $\mathcal{G}'$ recovers.

3.4. Implementation Details

In all experiments, we use four GNN layers, each outputting 256-dimensional embeddings. For the attention mechanism, we adopt 8 attention heads, aggregated by concatenation. We use AdamW as the optimizer with an initial learning rate of 5×10^{-4}. The hyperbolic curvature γ is initialized to -0.3. The margin λ in $\mathcal{L}_{\text{rec}}$ is set to 1.0, while α and β are tuned over $\{0.1, 0.5, 1.0\}$. Training continues up to 200 epochs on $4 \times$ NVIDIA V100 GPUs. For negative edges, we adopt a half random, half proximity-based sampling strategy, ensuring coverage of plausible but missing edges as well as obviously incorrect ones [47].

We also implement dynamic mini-batching: each batch includes a central node, a subset of its edges, and second-order neighbors. For each batch, we gather a random set of Horn clauses from local triads or short paths, thereby populating $\mathcal{L}_{\text{logic}}$. Gradients are accumulated across mini-batches before an update step to handle large graphs efficiently.

3.5. Quantitative Results

The Table summarizes the comparative performance. Our model consistently outperforms baseline methods across edge prediction, semantic consistency, and deductive closure metrics.

Method	Edge AUC	Sem. Consistency	Closure Rate
TransE	0.753	0.658	0.685
ComplEx	0.771	0.690	0.701
RGCN	0.781	0.704	0.729
GraphSAGE	0.776	0.695	0.712
KG-BERT	0.815	0.775	0.745
Ours	**0.923**	**0.892**	**0.942**

On ConceptNet 5.7, we achieve a 0.923 AUC, significantly higher than the next-best baseline (KG-BERT with 0.815). Semantic consistency ratings on newly inferred edges surpass all baselines by a comfortable margin (0.892 vs. 0.775 for KG-BERT). Notably, our deductive closure rate reaches 94.2%, compared to 68.5% for TransE, indicating the strength of our symbolic regularization. Similarly, performance on ATOMIC 2020 closely mirrors these trends.

3.6. Ablation Studies

We conduct ablation studies to systematically assess the contribution of each core model component. Our goal is to understand the sensitivity of the model to key architectural choices and regularization strategies by removing or modifying specific elements and observing the resulting performance degradation.

No Curvature Learning: In this setting, we replace the hyperbolic distance metric used in dynamic edge weighting with a Euclidean distance function, effectively constraining the representation space to a flat geometry. This modification results in a substantial 4.2% drop in edge AUC, underscoring the importance of curvature-aware representations for modeling hierarchical and transitive relationships. The performance degradation suggests that hyperbolic space provides a more natural embedding for structured knowledge graphs, particularly those exhibiting tree-like properties. By forcing the model to rely solely on Euclidean distances, the representation capacity for hierarchical relations is severely impaired, leading to suboptimal edge predictions [48] [49] [50].

No Deductive Regularization: We remove the logical consistency loss $\mathcal{L}_{\text{logic}}$, which enforces deductive constraints during training. Without this regularization, the model experiences a drastic reduction in closure rates, with values nearly halved across multiple reasoning benchmarks. This finding confirms that logic-based constraints play an essential role in ensuring coherent and transitive inference patterns. The lack of explicit logical supervision allows the model to generate more flexible but less reliable representations, often leading to violations of known logical rules. Our observations further indicate that while the model still maintains general semantic consistency, it struggles to enforce structured constraints such as transitivity and symmetry, which are crucial for robust knowledge inference.

No Attention Stream: We ablate the hierarchical attention mechanism, relying solely on edge-conditioned convolution to aggregate neighborhood information. This modification leads to a 6% drop in semantic consistency, indicating that attention mechanisms provide a crucial degree of flexibility in determining neighbor importance. Without attention-based weighting, the model assigns uniform significance to all neighboring nodes, which can lead to noisy or redundant information propagation. Our results show that attention-based aggregation helps refine neighborhood interactions, particularly in cases where certain edges contribute more significantly to logical consistency and predictive accuracy [51] [52] [53,54].

Our ablation results validate the necessity of three key components: (1) geometry-aware weighting through hyperbolic distance, (2) hierarchical attention for flexible neighbor aggregation, and (3) deductive constraints for enforcing logical structure. The interplay of these components enables our model to outperform baselines and maintain a structured representation of relational knowledge.

Ablation Setting	**Edge AUC Change (%)**	**Semantic Consistency Change (%)**
No Curvature Learning	-4.2%	-3.5%
No Deductive Regularization	-3.8%	-6.1%
No Attention Stream	-2.9%	-6.0%

Table 3. Performance impact of different ablation settings, showing drops in edge AUC and semantic consistency.

3.7. Qualitative Analysis

Beyond the quantitative evaluations, we analyze the interpretability and emergent structure of the learned representations. Our model demonstrates an ability to induce semantically meaningful clusters and hierarchies that reflect underlying conceptual relationships. In particular, we observe the following emergent properties:

Conceptual Layering: In commonsense graphs such as ConceptNet and ATOMIC, certain edge types, such as “causes” and “xEffect,” naturally form layered structures in hyperbolic space. For instance, a sequence of causal events such as “sleep deprivation” $\rightarrow$ “tiredness” $\rightarrow$ “accident risk” emerges as a structured trajectory, with each step following a logical progression. The hierarchical arrangement of concepts allows the model to capture transition dynamics more effectively, facilitating robust inference on unseen relations [55] [56].

Hierarchical Generalization: The “isA” relation forms deep hierarchical structures, with more general concepts positioned near the geometric center of hyperbolic space and more specific entities branching outward. This organization aligns with theoretical expectations, where broad categories such as “animal” or “vehicle” reside in central positions, while specific entities like "dog" or "car" extend further into space. The presence of such hierarchies supports the argument that hyperbolic embeddings inherently encode taxonomic structures in an optimal manner.

Contextual Refinement of Relations: One of the most intriguing phenomena observed in our analysis is the model's ability to refine ambiguous or noisy relations into more coherent categories. Specifically, in ConceptNet, edges labeled as "relatedTo" often serve as a catch-all for weakly associated concepts. However, our model learns to reassign some of these edges into more specific subcategories such as "similarTo" or "partOf" when sufficient contextual evidence is present. This emergent refinement suggests that the combination of attention-based weighting and deductive constraints helps the model resolve ambiguities in relational labeling [57] [58] [59].

Graph Structure Correction: Another emergent property of the model is its ability to collapse redundant or ill-defined edges. For example, in ATOMIC, the raw graph structure often contains multiple conflicting edges for the same node pair, leading to inconsistencies in reasoning tasks. Through joint optimization with logical constraints, our model naturally filters out implausible edges, effectively denoising the graph. This property enhances the reliability of downstream reasoning tasks, as the inferred structures align more closely with human intuition.

Semantic Consistency and Logical Coherence: The integration of hyperbolic geometry with deductive regularization not only enhances hierarchical organization but also strengthens logical coherence. Our qualitative evaluation suggests that the model learns to enforce transitivity and symmetry constraints implicitly, reducing inconsistencies in inferred relations. For example, if the model infers that "A isA B" and "B isA C," it is more likely to correctly infer "A isA C" without explicit supervision, demonstrating robust logical generalization.

To quantify these qualitative observations, we measure the model's ability to refine noisy relations and improve taxonomic structure. We introduce a metric called *Taxonomic Purity* ($\mathcal{P}_T$), which assesses the degree to which inferred hierarchies align with human-curated taxonomies. Higher $\mathcal{P}_T$ values indicate stronger semantic consistency in hierarchical embedding.

Model Variant	**Taxonomic Purity ($\mathcal{P}_T$)**	**Relation Refinement Rate (%)**
Full Model	85.7%	78.3%
No Deductive Regularization	74.5%	62.1%
No Attention Stream	76.2%	65.4%
Euclidean Space	67.9%	59.8%

Table 4. Taxonomic purity and relation refinement rate across different model configurations. The full model achieves the highest consistency, while Euclidean representations struggle to capture hierarchical structure.

Our analysis highlights that each component—hyperbolic representation, deductive constraints, and hierarchical attention—contributes significantly to both quantitative performance and qualitative interpretability. The model's ability to self-organize hierarchical structures, refine ambiguous relations, and correct graph inconsistencies demonstrates its potential for robust and explainable knowledge graph completion [60] [61] [62].

4. Theoretical Analysis

4.1. Expressiveness and Graph Isomorphism

We examine the capacity of our architecture to distinguish distinct graph structures, particularly in the presence of multi-relational edges. A well-known measure of expressiveness for GNNs is their equivalence (or lack thereof) to the Weisfeiler-Lehman (WL) test.

Theorem 1

The proposed hybrid GNN architecture is at least as powerful as the 1-WL test in distinguishing non-isomorphic multi-relational graphs.

Proof Sketch: Our edge-conditioned convolution captures relation-specific neighbor aggregation, which generalizes the 1-WL color refinement on node labels. The hierarchical attention mechanism further introduces a non-uniform weighting scheme over neighbors. By combining these two aspects, the model can differentiate between nodes that would otherwise be considered equivalent under standard message passing. Moreover, the presence of an explicit relational encoding, $\mathbf{M}_r$, ensures that different relation types do not collapse into identical aggregation patterns, thereby augmenting WL's node label refinements [63] [64] [65]. In simpler terms, if two multi-relational graphs are distinct, they will eventually produce different embedding distributions under our proposed architecture, provided sufficient layers and appropriate parameterization. Thus, we do not lose expressiveness compared to standard relational GNN approaches; we strictly gain from attention and curvature-based expansions.

4.2. Hyperbolic Geometry for Hierarchical Data

Hierarchically structured data often violates Euclidean distance properties, since the distance between siblings (e.g., "dog" and "cat") might appear small, but their common ancestor in the "mammal" or "animal" category might be conceptually close to one or both in a way that Euclidean geometry cannot easily capture without significant dimensional expansion. Hyperbolic or Lorentzian spaces, by contrast, can embed trees or tree-like structures with far lower distortion [66] [67].

When $\gamma < 0$, the model primarily gains from the exponential volume growth of hyperbolic space, enabling nodes deeper in the hierarchy to reside further from a conceptual center. Our approach learns γ to adjust curvature such that the network can discover an optimal embedding manifold. If higher-level relations need more hierarchical structure, γ moves to more negative values, intensifying hyperbolic effects. Conversely, if data exhibits less hierarchy, γ shifts closer to 0, approaching a near-Euclidean regime.

4.3. Logical Constraints and Consistency

Symbolic constraints enforced by $\mathcal{L}_{\text{logic}}$ ensure that subgraphs form consistent sets of implications. Specifically, we consider Horn clauses of the form:

$$(r, t \to s) \quad \Leftrightarrow \quad e^s_{ik} \leftarrow e^r_{ij} \wedge e^t_{jk}.$$

We unify these clauses in a differentiable manner by penalizing large discrepancies between $\phi(e^s_{ik})$ and $\min(\phi(e^r_{ij}), \phi(e^t_{jk}))$. This approach is more flexible than strictly enforcing symbolic constraints, allowing the model to weigh the cost of potential clause violations against the overall reconstruction objectives. The net result is a balance that yields high deductive coverage without forcing every possible transitive edge to appear explicitly.

Moreover, to maintain *consistency*, we can introduce additional constraints for mutually exclusive relations or contradictory statements. For instance, if a domain rule states that "an entity cannot simultaneously be a dog and a fish," the model can incorporate a pairwise penalty whenever $\phi(e_{iv}^{\text{isA=dog}})$ and $\phi(e_{iv}^{\text{isA=fish}})$ are both high.

5. Conclusion

We have introduced a novel framework that addresses the crucial challenges of scaling, pruning, and logically structuring commonsense knowledge bases. Our solution combines (1) an edge-conditioned GNN stream that handles multi-relational data by leveraging relation-specific transformations, (2) a hierarchical attention mechanism that dynamically refines or prunes edges, (3) curvature-aware embedding in hyperbolic or Lorentzian spaces to capture hierarchical relations, and (4) a multi-task learning objective that balances reconstruction fidelity, deductive consistency, and topological sparsity.

Empirical results on ConceptNet and ATOMIC confirm significant improvements over existing baselines such as TransE, ComplEx, and GNN variants. Our method not only achieves superior edge prediction performance but also excels in deductive closure, confirming it systematically identifies transitive and causal links. Ablation studies demonstrate that removing curvature learning or logical constraints materially degrades performance, underscoring the effectiveness of integrating geometry and symbolic logic [68] [69] [70].

From a theoretical standpoint, we show that our hybrid architecture is at least as expressive as the Weisfeiler-Lehman test, even in multi-relational contexts, and that hyperbolic geometry provides a powerful inductive bias for hierarchical or tree-like structures often found in commonsense domains. The differentiable logic constraints allow the model to elegantly reconcile structural completeness with interpretability, avoiding an explosion of spurious or contradictory edges.

Our work opens several avenues for future exploration. First, extending the approach to *temporal commonsense* knowledge could capture evolving relations (e.g., "morning routine" changes to "afternoon routine"). Second, integrating large language models as external knowledge sources or as means to generate probable Horn clauses could further improve coverage. Third, investigating advanced meta-reasoning tasks, where the model must reconcile contradictory statements or weigh the reliability of different sources, stands as another promising direction [71] [72] [73]. Finally, the techniques developed here, particularly those involving joint geometric-symbolic optimization, may prove relevant to structuring broader classes of knowledge graphs beyond commonsense, including specialized scientific or biomedical ontologies. Uniting the strengths of geometric representation learning, attention-based relational modeling, and differentiable logic constraints, our framework paves the way for more coherent, scalable, and interpretable commonsense knowledge bases. We envision this methodology as a step toward advanced AI systems capable of flexible, robust reasoning across diverse domains of everyday human experience [74] [75] [76].

References

1. Zerilli, J.; Knott, A.; Maclaurin, J.; Gavaghan, C. Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology* **2018**, *32*, 661–683. https://doi.org/10.1007/s13347-018-0330-6.
2. Koriche, F. Learning to assign degrees of belief in relational domains. *Machine Learning* **2008**, *73*, 25–53. https://doi.org/10.1007/s10994-008-5075-5.
3. Bannour, H.; Hudelot, C. Building and using fuzzy multimedia ontologies for semantic image annotation. *Multimedia Tools and Applications* **2013**, *72*, 2107–2141. https://doi.org/10.1007/s11042-013-1491-z.
4. Hooker, C.A. Interaction and bio-cognitive order. *Synthese* **2008**, *166*, 513–546. https://doi.org/10.1007/s11229-008-9374-y.
5. Hoehndorf, R.; Loebe, F.; Kelso, J.; Herre, H. Representing default knowledge in biomedical ontologies: application to the integration of anatomy and phenotype ontologies. *BMC bioinformatics* **2007**, *8*, 377–377. https://doi.org/10.1186/1471-2105-8-377.
6. Eiter, T.; Erdem, E.; Fink, M.; Senko, J. Comparing action descriptions based on semantic preferences. *Annals of Mathematics and Artificial Intelligence* **2007**, *50*, 273–304. https://doi.org/10.1007/s10472-007-9077-y.
7. O'Shea, J. The 'theory theory' of mind and the aims of Sellars' original myth of Jones. *Phenomenology and the Cognitive Sciences* **2012**, *11*, 175–204. https://doi.org/10.1007/s11097-011-9250-y.
8. Miyahara, K. Neo-pragmatic intentionality and enactive perception: a compromise between extended and enactive minds. *Phenomenology and the Cognitive Sciences* **2011**, *10*, 499–519. https://doi.org/10.1007/s11097-011-9212-4.
9. Wheeldon, J.; Heidt, J. Bridging the Gap: A Pragmatic Approach to Understanding Critical Criminologies and Policy Influence. *Critical Criminology* **2007**, *15*, 313–325. https://doi.org/10.1007/s10612-007-9041-5.
10. Sharma, A.; Forbus, K.D. Graph-based reasoning and reinforcement learning for improving Q/A performance in large knowledge-based systems. In Proceedings of the 2010 AAAI Fall Symposium Series, 2010.

11. Sammel, A.J. Traveling with and through your backpack: a personal reflection on the infrastructure of science education. *Cultural Studies of Science Education* **2008**, *3*, 843–857. https://doi.org/10.1007/s11422-008-9118-9.
12. Urbaniak, R. Narration in judiciary fact-finding: a probabilistic explication. *Artificial Intelligence and Law* **2018**, *26*, 345–376. https://doi.org/10.1007/s10506-018-9219-z.
13. Graham, P.J. Formulating reductionism about testimonial warrant and the challenge from childhood testimony. *Synthese* **2016**, *195*, 3013–3033. https://doi.org/10.1007/s11229-016-1140-y.
14. Davies, W. The inscrutability of colour similarity. *Philosophical Studies* **2014**, *171*, 289–311. https://doi.org/10.1007/s11098-013-0272-x.
15. Lehmann, J.; Gangemi, A. An ontology of physical causation as a basis for assessing causation in fact and attributing legal responsibility. *Artificial Intelligence and Law* **2007**, *15*, 301–321. https://doi.org/10.1007/s10506-007-9035-3.
16. Ikuenobe, P. The Practical and Experiential Reality of Racism: Carter's and Corlett's Realism About Race and Racism. *Journal of African American Studies* **2018**, *22*, 373–392. https://doi.org/10.1007/s12111-018-9417-5.
17. Hoffmann, G. Truth, Superassertability, and Conceivability. *The Journal of Value Inquiry* **2008**, *42*, 287–299. https://doi.org/10.1007/s10790-008-9125-9.
18. Sharma, A.; Goolsbey, K. Identifying useful inference paths in large commonsense knowledge bases by retrograde analysis. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2017, Vol. 31.
19. Wang, Y. Automatic semantic analysis of software requirements through machine learning and ontology approach. *Journal of Shanghai Jiaotong University (Science)* **2016**, *21*, 692–701. https://doi.org/10.1007/s12204-016-1783-3.
20. Emirbayer, M.; Noble, M. The peculiar convergence of Jeffrey Alexander and Erik Olin Wright. *Theory and Society* **2013**, *42*, 617–645. https://doi.org/10.1007/s11186-013-9201-4.
21. Nyga, D.; Beetz, M., ISRR (2) - Cloud-based Probabilistic Knowledge Services for Instruction Interpretation; Springer International Publishing, 2017; pp. 649–664. https://doi.org/10.1007/978-3-319-60916-4_37.
22. Sakama, C.; Inoue, K. Inductive equivalence in clausal logic and nonmonotonic logic programming. *Machine Learning* **2010**, *83*, 1–29. https://doi.org/10.1007/s10994-010-5189-4.
23. Eede, Y.V.D. The (Im)Possible Grasp of Networked Realities: Disclosing Gregory Bateson's Work for the Study of Technology. *Human Studies* **2016**, *39*, 601–620. https://doi.org/10.1007/s10746-016-9400-x.
24. Garofalo, G.; Fetoni, P. The Chicago School after the crisis of the new millennium. *Quality & Quantity* **2011**, *47*, 677–711. https://doi.org/10.1007/s11135-011-9539-5.
25. Aakur, S.N.; de Souza, F.D.; Sarkar, S. Going Deeper with Semantics: Video Activity Interpretation using Semantic Contextualization., 2017.
26. Laszlo, P. Towards Teaching Chemistry as a Language. *Science & Education* **2011**, *22*, 1669–1706. https://doi.org/10.1007/s11191-011-9408-6.
27. *AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning - Symbolic Probabilistic Reasoning for Narratives*, 2011.
28. He, L.; Liu, B.; Li, G.; Sheng, Y.; Wang, Y.; Xu, Z. Knowledge Base Completion by Variational Bayesian Neural Tensor Decomposition. *Cognitive Computation* **2018**, *10*, 1075–1084. https://doi.org/10.1007/s12559-018-9565-x.
29. Shum, H.Y.; He, X.; Li, D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* **2018**, *19*, 10–26. https://doi.org/10.1631/fitee.1700826.
30. Tavassoli, N.T. So you Think You Know your Customers. *International Commerce Review* **2011**, *10*, 56–69. https://doi.org/10.1007/s12146-011-0067-y.
31. Sharma, A. Structural and network-based methods for knowledge-based systems. PhD thesis, Northwestern University, 2011.
32. van Riemsdijk, M.B.; Dastani, M.; Meyer, J.J.C. Goals in conflict: semantic foundations of goals in agent programming. *Autonomous Agents and Multi-Agent Systems* **2008**, *18*, 471–500. https://doi.org/10.1007/s10458-008-9067-4.
33. Ming, T. Who Does the Sounding? The Metaphysics of the First-Person Pronoun in the Zhuangzi. *Dao* **2016**, *15*, 57–79. https://doi.org/10.1007/s11712-015-9474-6.

34. Lemaignan, S.; Ros, R.; Sisbot, E.A.; Alami, R.; Beetz, M. Grounding the Interaction: Anchoring Situated Discourse in Everyday Human-Robot Interaction. *International Journal of Social Robotics* **2011**, *4*, 181–199. https://doi.org/10.1007/s12369-011-0123-x.
35. Bollen, K.; Ittner, H.; Euwema, M. Mediating Hierarchical Labor Conflicts: Procedural Justice Makes a Difference—for Subordinates. *Group Decision and Negotiation* **2011**, *21*, 621–636. https://doi.org/10.1007/s10726-011-9230-1.
36. Maynard, D.W.; Turowetz, J. Doing Testing: How Concrete Competence can Facilitate or Inhibit Performances of Children with Autism Spectrum Disorder. *Qualitative Sociology* **2017**, *40*, 467–491. https://doi.org/10.1007/s11133-017-9368-5.
37. Lee, G.; Byun, T. An Explanation for the Difficulty of Leading Conceptual Change Using a Counterintuitive Demonstration: The Relationship Between Cognitive Conflict and Responses. *Research in Science Education* **2011**, *42*, 943–965. https://doi.org/10.1007/s11165-011-9234-5.
38. Doan, A.; Madhavan, J.; Dhamankar, R.; Domingos, P.; Halevy, A. Learning to match ontologies on the Semantic Web. *The VLDB Journal The International Journal on Very Large Data Bases* **2003**, *12*, 303–319. https://doi.org/10.1007/s00778-003-0104-2.
39. Wang, Y.; Li, Z.; Tang, Z.; Zeng, G. A GIS-Based Spatial Multi-Criteria Approach for Flood Risk Assessment in the Dongting Lake Region, Hunan, Central China. *Water Resources Management* **2011**, *25*, 3465–3484. https://doi.org/10.1007/s11269-011-9866-2.
40. Haotian, X.; Peng, H.; Xie, H.; Cambria, E.; Liuyang, Z.; Weiguo, Z. End-to-End latent-variable task-oriented dialogue system with exact log-likelihood optimization. *World Wide Web* **2019**, *23*, 1989–2002. https://doi.org/10.1007/s11280-019-00688-8.
41. Sharma, A.; Forbus, K. Automatic extraction of efficient axiom sets from large knowledge bases. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2013, Vol. 27, pp. 1248–1254. https://doi.org/10.1609/aaai.v27i1.8472.
42. Miščević, N. No More Tears in Heaven: Two Views of Response-Dependence. *Acta Analytica* **2011**, *26*, 75–93. https://doi.org/10.1007/s12136-010-0121-x.
43. Moghimi, R.; Anvari, A. An integrated fuzzy MCDM approach and analysis to evaluate the financial performance of Iranian cement companies. *The International Journal of Advanced Manufacturing Technology* **2013**, *71*, 685–698. https://doi.org/10.1007/s00170-013-5370-6.
44. Jones, R.C. Science, sentience, and animal welfare. *Biology & Philosophy* **2012**, *28*, 1–30. https://doi.org/10.1007/s10539-012-9351-1.
45. Coleman, S. The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence. *Erkenntnis* **2013**, *79*, 19–44. https://doi.org/10.1007/s10670-013-9431-x.
46. Borgwardt, S.; Ceylan, I.I.; Lukasiewicz, T. AAAI - OntologyMediated Query Answering over LogLinear Probabilistic Data. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**, *33*, 2711–2718. https://doi.org/10.1609/aaai.v33i01.33012711.
47. Sharma, A.; Forbus, K. Graph traversal methods for reasoning in large knowledge-based systems. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2013, Vol. 27, pp. 1255–1261.
48. Maclean, K. Re-conceptualising desert landscapes: unpacking historical narratives and contemporary realities for sustainable livelihood development in central Australia. *GeoJournal* **2008**, *74*, 451–463. https://doi.org/10.1007/s10708-008-9234-9.
49. Furbach, U.; Hölldobler, S.; Ragni, M.; Schon, C.; Stolzenburg, F. Cognitive Reasoning: A Personal View. *KI - Künstliche Intelligenz* **2019**, *33*, 209–217. https://doi.org/10.1007/s13218-019-00603-3.
50. Saba, W.S. Commonsense Knowledge, Ontology and Ordinary Language, 2008.
51. Ragni, M.; Eichhorn, C.; Bock, T.; Kern-Isberner, G.; Tse, A.P. Formal Nonmonotonic Theories and Properties of Human Defeasible Reasoning. *Minds and Machines* **2017**, *27*, 79–117. https://doi.org/10.1007/s11023-016-9414-1.
52. Pigozzi, G.; Tsoukiàs, A.; Viappiani, P. Preferences in artificial intelligence. *Annals of Mathematics and Artificial Intelligence* **2015**, *77*, 361–401. https://doi.org/10.1007/s10472-015-9475-5.
53. Li, J. IMCOM - Domain specific commonsense relation extraction from bag of concepts metadata. In Proceedings of the Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication. ACM, January 2015, pp. 96–4. https://doi.org/10.1145/2701126.2701159.
54. Sharma, A.; Goolsbey, K.M. Simulation-based approach to efficient commonsense reasoning in very large knowledge bases. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2019, Vol. 33, pp. 1360–1367.

55. Pietsch, W. A Causal Approach to Analogy. *Journal for General Philosophy of Science* **2019**, *50*, 489–520. https://doi.org/10.1007/s10838-019-09463-9.
56. Lin, X. People's Justice: Socialist Law and Equity in China, 1921–1945. *Fudan Journal of the Humanities and Social Sciences* **2019**, *12*, 473–491. https://doi.org/10.1007/s40647-019-00264-4.
57. von Furstenberg, G.M.; Jeong, J.H. Owning Up to Uncertainty in Macroeconomics. *The Geneva Papers on Risk and Insurance - Issues and Practice* **1988**, *13*, 12–90. https://doi.org/10.1057/gpp.1988.2.
58. *AAAI/IAAI - Elaboration tolerance of logical theories*, 1999.
59. Bello, P.; Bringsjord, S. On How to Build a Moral Machine. *Topoi* **2012**, *32*, 251–266. https://doi.org/10.1007/s11245-012-9129-8.
60. Elstein, D. Why Early Confucianism Cannot Generate Democracy. *Dao* **2010**, *9*, 427–443. https://doi.org/10.1007/s11712-010-9187-9.
61. Müller, T. Branching in the landscape of possibilities. *Synthese* **2012**, *188*, 41–65. https://doi.org/10.1007/s11229-011-0059-6.
62. Raibley, J. Health and well-being. *Philosophical Studies* **2012**, *165*, 469–489. https://doi.org/10.1007/s11098-012-9951-2.
63. Tran, H.N.; Cambria, E. Ensemble application of ELM and GPU for real-time multimodal sentiment analysis. *Memetic Computing* **2017**, *10*, 3–13. https://doi.org/10.1007/s12293-017-0228-3.
64. Malik, K.R.; Ahmad, T.; Farhan, M.; Aslam, M.; Jabbar, S.; Khalid, S.; Kim, M. Big-data: transformation from heterogeneous data to semantically-enriched simplified data. *Multimedia Tools and Applications* **2015**, *75*, 12727–12747. https://doi.org/10.1007/s11042-015-2918-5.
65. Kaniklidis, C.; Mezei, L.; Kittredge, R.; Parker, A.C.E.; Cline, H.F.; Nuessel, F.H.; Beshers, J.M.; Clark, G.A.; Hall, J.F.; Findler, C.; et al. Book reviews. *Computers and the Humanities* **1979**, *13*, 311–333. https://doi.org/10.1007/bf02400144.
66. Hertzberg, J.; Zhang, J.; Zhang, L.; Rockel, S.; Neumann, B.; Lehmann, J.; Dubba, K.; Cohn, A.G.; Saffiotti, A.; Pecora, F.; et al. The RACE Project : Robustness by Autonomous Competence Enhancement. *KI - Künstliche Intelligenz* **2014**, *28*, 297–304. https://doi.org/10.1007/s13218-014-0327-y.
67. Han, J.; Cheng, H.; Xin, D.; Yan, X. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery* **2007**, *15*, 55–86. https://doi.org/10.1007/s10618-006-0059-1.
68. Kutateladze, S.S. Multiobjective problems of convex geometry. *Siberian Mathematical Journal* **2009**, *50*, 887–897. https://doi.org/10.1007/s11202-009-0099-z.
69. Foo, N.; Low, B.T. A Note on Prototypes, Convexity and Fuzzy Sets. *Studia Logica* **2008**, *90*, 125–137. https://doi.org/10.1007/s11225-008-9146-1.
70. Sutcliffe, G. The TPTP Problem Library and Associated Infrastructure. *Journal of Automated Reasoning* **2009**, *43*, 337–362. https://doi.org/10.1007/s10817-009-9143-8.
71. Darder, A.; Torres, R.D. Mapping Latino Studies: Critical Reflections on Class and Social Theory. *Latino Studies* **2003**, *1*, 303–324. https://doi.org/10.1057/palgrave.lst.8600027.
72. Sakama, C.; Inoue, K. Brave induction: a logical framework for learning from incomplete information. *Machine Learning* **2009**, *76*, 3–35. https://doi.org/10.1007/s10994-009-5113-y.
73. Rivero-Obra, M. Dramatizing The SUBJECT'S Identity. *Philosophia* **2018**, *47*, 1227–1245. https://doi.org/10.1007/s11406-018-0023-5.
74. Huang, X.; Huang, H.; Liao, B.; Xu, C. An Ontology-Based Approach to Metaphor Cognitive Computation. *Minds and Machines* **2012**, *23*, 105–121. https://doi.org/10.1007/s11023-012-9269-z.
75. Nakatani, C.; Chehelcheraghi, M.; Jarrahi, B.; Nakatani, H.; van Leeuwen, C. Perceivers' internal state tags fixation-by-fixation visual information: An EEG-eye movement co-registration study. *Cognitive Processing* **2015**, *16*, 99–99. https://doi.org/10.1007/s10339-015-0732-7.
76. Macdonald, I. On the 'undialectical': normativity in Hegel. *Continental Philosophy Review* **2011**, *45*, 121–141. https://doi.org/10.1007/s11007-011-9211-8.