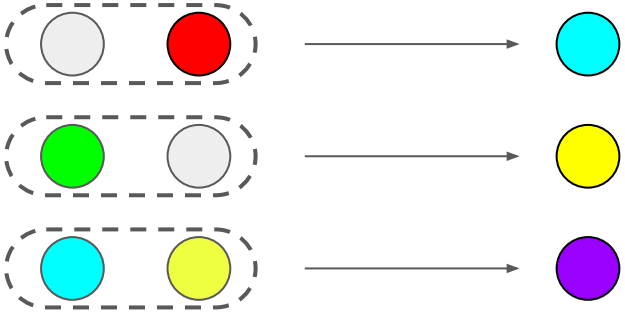
Systematic relational reasoning using Epistemic Graph Neural Networks

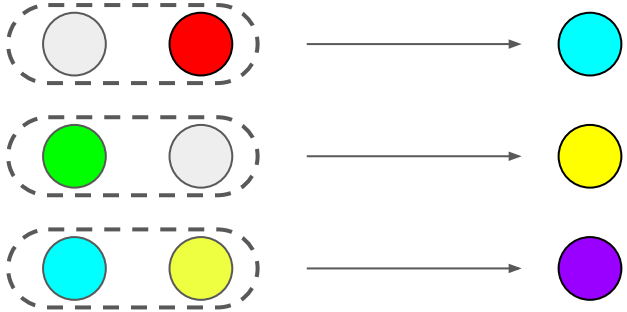
Irtaza Khalid, Steven Schockaert

What is systematic relational reasoning and why is it important?

What is Systematic relational reasoning?

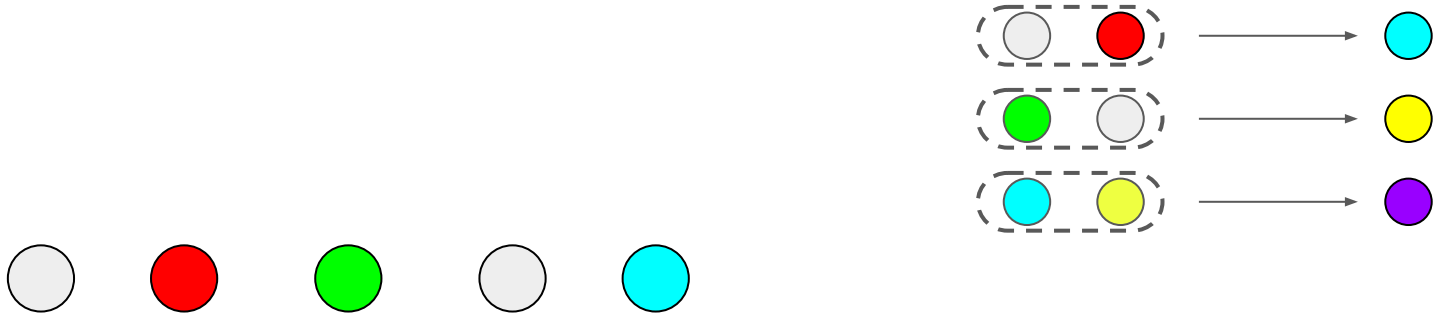


What is Systematic relational reasoning?



given base compositions

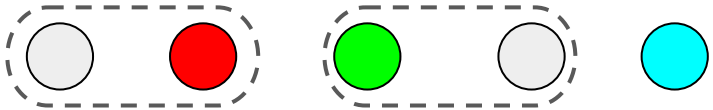
What is Systematic relational reasoning?

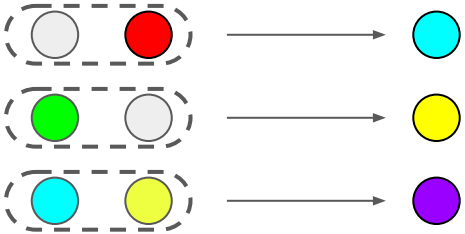


evaluate the ability of a machine to recombine base compositions to collapse increasingly long sequences

Hupkes et. al. 2020

What is Systematic relational reasoning?

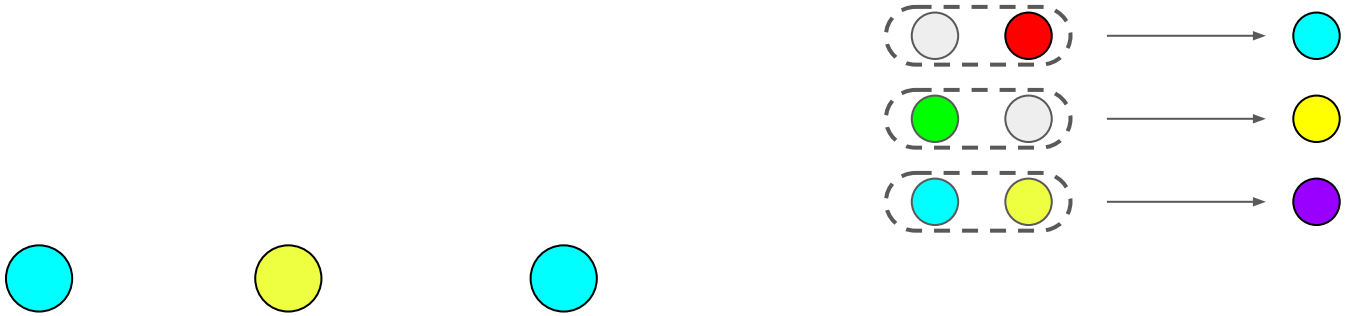




evaluate the ability of a machine to recombine base compositions to collapse increasingly long sequences

Hupkes et. al. 2018

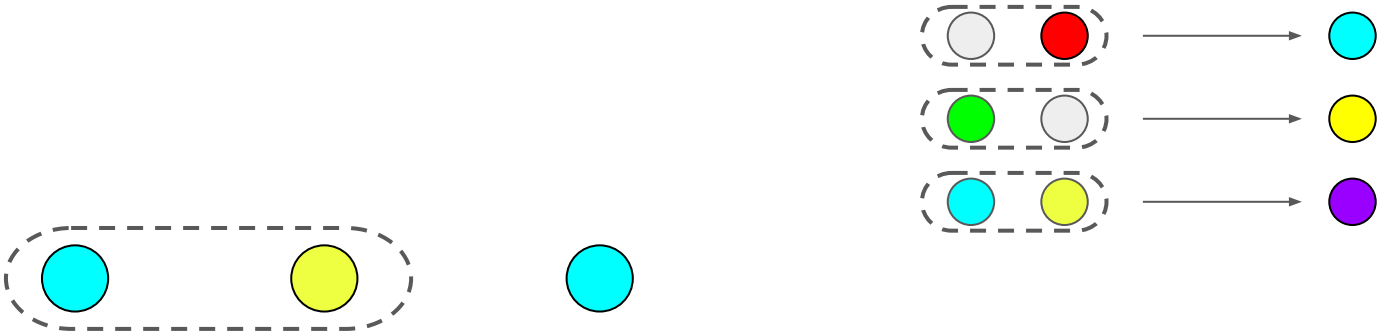
What is Systematic relational reasoning?



evaluate the ability of a machine to recombine base compositions to collapse increasingly long sequences

Hupkes et. al. 2020

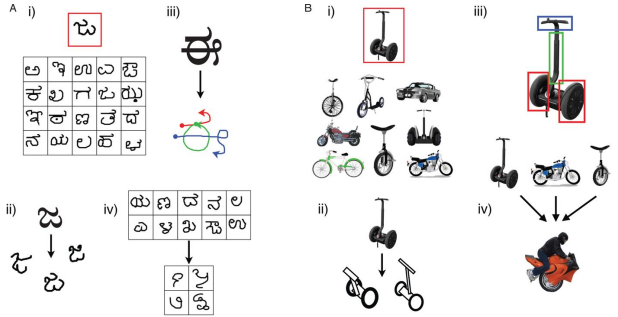
What is Systematic relational reasoning?



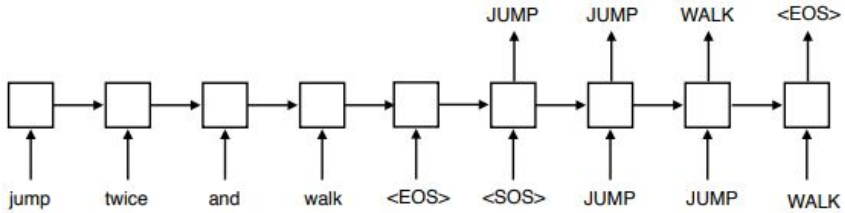
evaluate the ability of a machine to recombine base compositions to collapse increasingly long sequences

Hupkes et. al. 2020

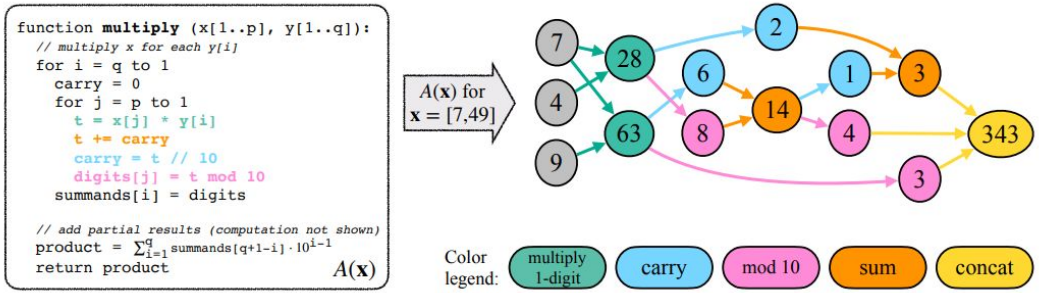
Compositional learning unlocks human-like learning with sparse data



Lake et. al. 2017



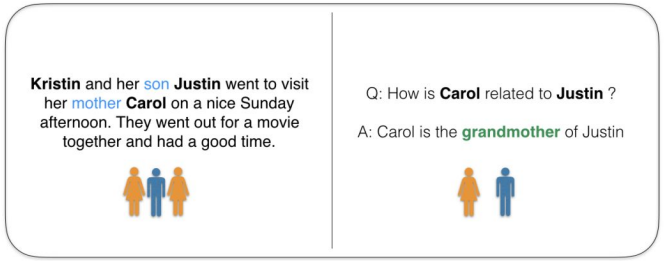
Lake et. al. 2018

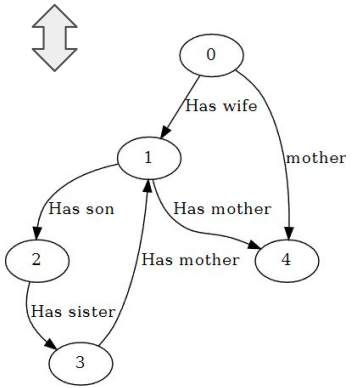


Dziri et. al. 2023

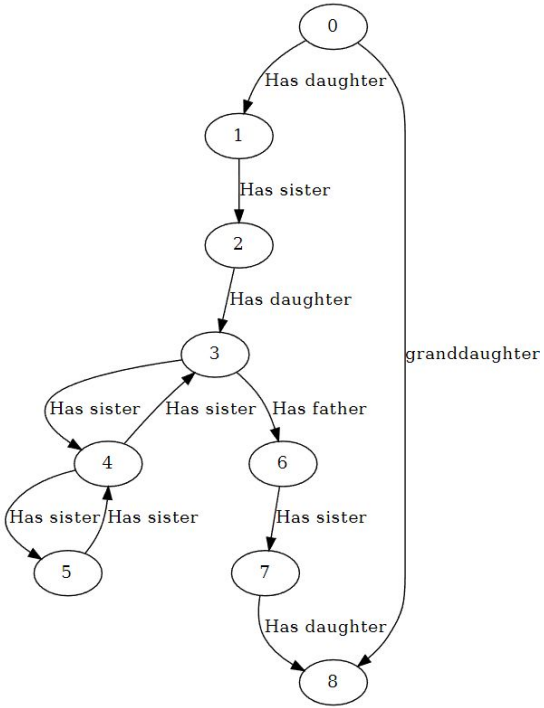
How to measure systematic *relational* reasoning?
⇒ link prediction problem: (s,?,t)

Train on small graphs
(path length from source to sink: k=2,3,4)





Test on increasingly large graphs
(path length k=2,3,4,5,...,10)

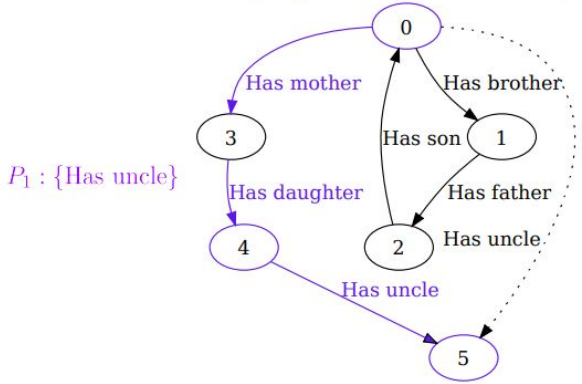


CLUTRR (Sinha et. al. 2022)

Limitations of current methods

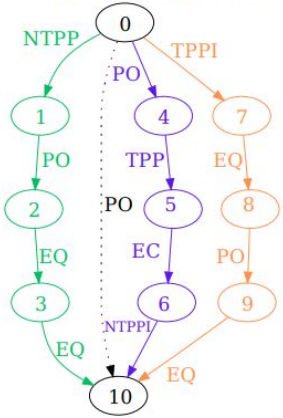
1. Though neuro-symbolic (NeSy) or neural-theorem-prover type methods are good at these problems,
    a. They are too specialized for single path conjunctive reasoning
    b. They are parameter-inefficient and generally not very scalable to large graphs
2. Non-NeSy *statistical* methods are generally poor at systematic reasoning. We argue that is because
    a. They lack an algorithmical alignment bias (Xu et. al. 2021) wrt. the systematicity task
    b. They are prone to learning shortcuts (spurious patterns) (Gierhos et. al.)

Single-path reasoning

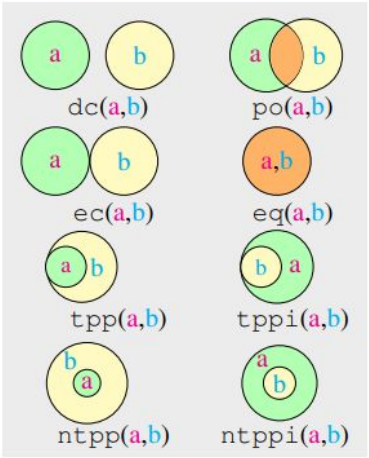


$$r_3(X, Z) \leftarrow r_1(X, Y) \wedge r_2(Y, Z)$$

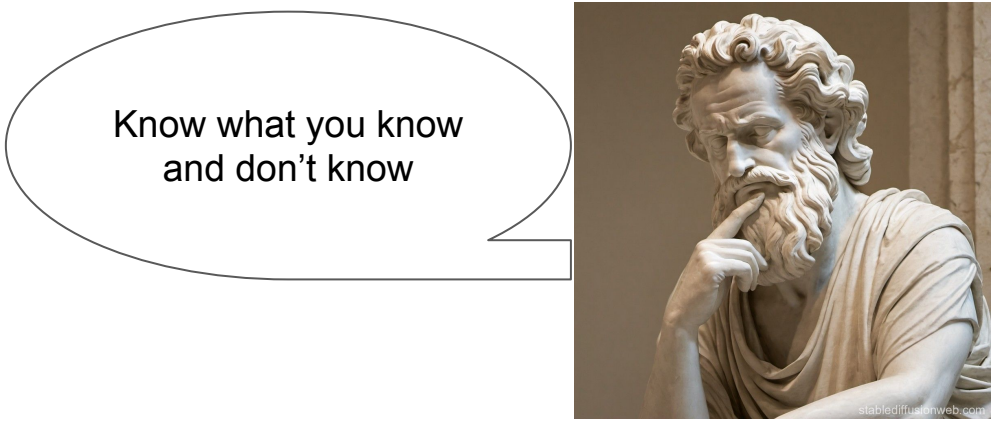
Multi-path disjunctive reasoning



P_1 : {DC, EC, PO, TPP, NTPP}
P_2 : {NTPPI, DC, EC, TPPI, PO}
P_3 : {TPPI, NTPPI, PO}
$P_1 \cap P_2 \cap P_3$: {PO}



$$s_1(X, Z) \vee \ldots \vee s_k(X, Z) \leftarrow r_1(X, Y) \wedge r_2(Y, Z)$$
$$(s_1 \vee \cdots \vee s_k)_{\text{path}_1} \wedge \cdots \wedge (s_1 \vee \cdots \vee s_k)_{\text{path}_b}$$



EpiGNN: the Epistemic Graph Neural Network

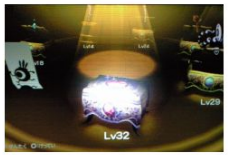
Idea: Algorithmically aligning the architecture with the algorithm that solves the task aids generalization (Xu et. al. 2020)

Graph Neural Network	Bellman-Ford algorithm
for k = 1 … GNN iter:	for k = 1 … |S| - 1:
for u in S:	for u in S:
$h_u^{(k)} = \Sigma_v \text{ MLP}(h_v^{(k-1)}, h_u^{(k-1)})$	$d[k][u] = \min_v d[k-1][v] + \text{cost}(v, u)$

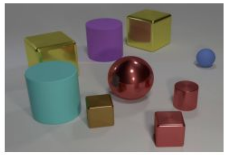
No need to learn for-loops

Learns a simple reasoning step

MLPs have to learn for-loops that GNNs don't so tasks unified by dynamic programming are more sample efficiently learned

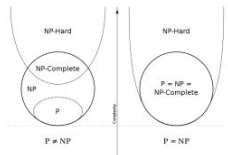


Summary statistics
What is the maximum value difference among treasures?



Relational argmax
What are the colors of the furthest pair of objects?

Dynamic programming
What is the cost to defeat monster X by following the optimal path?



NP-hard problem
Subset sum: Is there a subset that sums to 0?

So, algorithmically align the GNN with the algebraic closure algorithm that “solves” the general reasoning problem

Use NBFNet for relational embeddings:

1. Initialize probabilistic embeddings that can encode unions of base relations
2. Use a COMBINE function that simulates discrete relational composition.
3. Use an AGGREGATE function that simulates intersection.
4. Repeat for k rounds.

Main steps in algebraic closure are:

1. Initialize node embeddings for all possible relations.
2. Compute all possible discrete relational compositions.
3. Update node embeddings via intersection with the set from step 2.
4. Repeat for k iterations.

Detailed: Neural vs. Classical

1. Initialization

$$\mathbf{e}^{(0)} = \begin{cases} (1, 0, \ldots, 0) & \text{if } e = h \\ (\frac{1}{n}, \ldots, \frac{1}{n}) & \text{otherwise} \end{cases}$$

$$X_{ef}^{(0)} = \begin{cases} \{r\} & \text{if } r(e,f) \in \mathcal{F} \\ \{\hat{r}\} & \text{if } r(f,e) \in \mathcal{F} \\ \mathcal{R} & \text{otherwise} \end{cases}$$

2. Message passing (COMBINE) composition

$$\phi((f_1, \ldots, f_n), (r_1, \ldots, r_n)) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i r_j \mathbf{a}_{ij}$$

$$X_{eg}^{(i-1)} \diamond X_{gf}^{(i-1)} = \bigcup \left\{ r \circ s | r \in X_{eg}^{(i-1)}, s \in X_{gf}^{(i-1)} \right\}$$

3. Aggregation: ψ `min` or `mul`

$$\mathbf{e}^{(l)} = \psi(\{\mathbf{e}^{(l-1)}\} \cup \{\phi(\mathbf{r}, \mathbf{f}^{(l-1)}) \,|\, r(e,f) \in \mathcal{F}\})$$

$$X_{ef}^{(i)} = X_{ef}^{(i-1)} \cap \bigcap \{X_{eg}^{(i-1)} \diamond X_{gf}^{(i-1)} \,|\, g \in \mathcal{E}\}$$

Results

Link prediction (h,?,t): CLUTRR

Table 1: Results (accuracy) on CLUTRR after training on problems with $k \in \{2, 3, 4\}$ and then evaluating on problems with $k \in \{5, \ldots, 10\}$. Results marked with * were taken from (Minervini et al., 2020b), those with † from (Lu et al., 2022) and those with [2] from (Cheng et al., 2023). The best performance for each k is highlighted in **bold**.

	5 Hops	6 Hops	7 Hops	8 Hops	9 Hops	10 Hops
EpiGNN-mul (ours)	0.99±.01	**0.99±.01**	**0.99±.02**	0.99±.03	0.96±.03	**0.98±.02**
EpiGNN-min (ours)	0.99±.01	0.98±.02	0.98±.03	0.97±.06	0.95±.04	0.93±.07
NCRL[2]	**1.0±.01**	**0.99±.01**	0.98±.02	0.98±.03	0.98±.03	0.97±.02
R5†	0.99±.02	0.99±.04	0.99±.03	**1.0±.02**	**0.99±.02**	0.98±.03
CTP^*_L	0.99±.02	0.98±.04	0.97±.04	0.98±.03	0.97±.04	0.95±.04
CTP^*_A	0.99±.04	0.99±.03	0.97±.03	0.95±.06	0.93±.07	0.91±.05
CTP^*_M	0.98±.04	0.97±.06	0.95±.06	0.94±.08	0.93±.08	0.90±.09
GNTP*	0.68±.28	0.63±.34	0.62±.31	0.59±.32	0.57±.34	0.52±.32
ET	0.99±.01	0.98±.02	**0.99±.02**	0.96±.04	0.92±.07	0.92±.07
GAT*	0.99±.00	0.85±.04	0.80±.03	0.71±.03	0.70±.03	0.68±.02
GCN*	0.94±.03	0.79±.02	0.61±.03	0.53±.04	0.53±.04	0.41±.04
NBFNet	0.83±.11	0.68±.09	0.58±.10	0.53±.07	0.50±.11	0.53±.08
R-GCN	0.97±.03	0.82±.11	0.60±.13	0.52±.11	0.50±.09	0.45±.09
RNN*	0.93±.06	0.87±.07	0.79±.11	0.73±.12	0.65±.16	0.64±.16
LSTM*	0.98±.03	0.95±.04	0.89±.10	0.84±.07	0.77±.11	0.78±.11
GRU*	0.95±.04	0.94±.03	0.87±.08	0.81±.13	0.74±.15	0.75±.15

NeSy

GNNs

Link prediction (h,?,t): GraphLOG

Table 2: Results on Graphlog (accuracy). For each world, we report the number of distinct relation sequences between head and tail (ND) and the Average resolution length (ARL). Results marked with * were taken from (Lu et al., 2022) and those with † from (Cheng et al., 2023). The best and second-best performance across all the models are highlighted in **bold** or underlined.

World ID	ND	ARL	E-GAT*	R-GCN*	CTP*	R5*	NCRL†	ET	EpiGNN-mul
World 6	249	5.06	0.536	0.498	0.533±0.03	0.687±0.05	**0.702±0.02**	0.496 ± 0.087	0.648 ± 0.012
World 7	288	4.47	0.613	0.537	0.513±0.03	**0.749±0.04**	-	0.487 ± 0.056	0.611±0.026
World 8	404	5.43	0.643	0.569	0.545±0.02	0.671±0.03	**0.687**±0.02	0.55 ± 0.092	0.649±0.042
World 11	194	4.29	0.552	0.456	0.553±0.01	**0.803**±0.01	-	0.637 ± 0.091	0.758 ± 0.037
World 32	287	4.66	0.700	0.621	0.581±0.04	0.841±0.03	-	0.815 ± 0.061	**0.914±0.026**

Link prediction (h,?,t): STaR

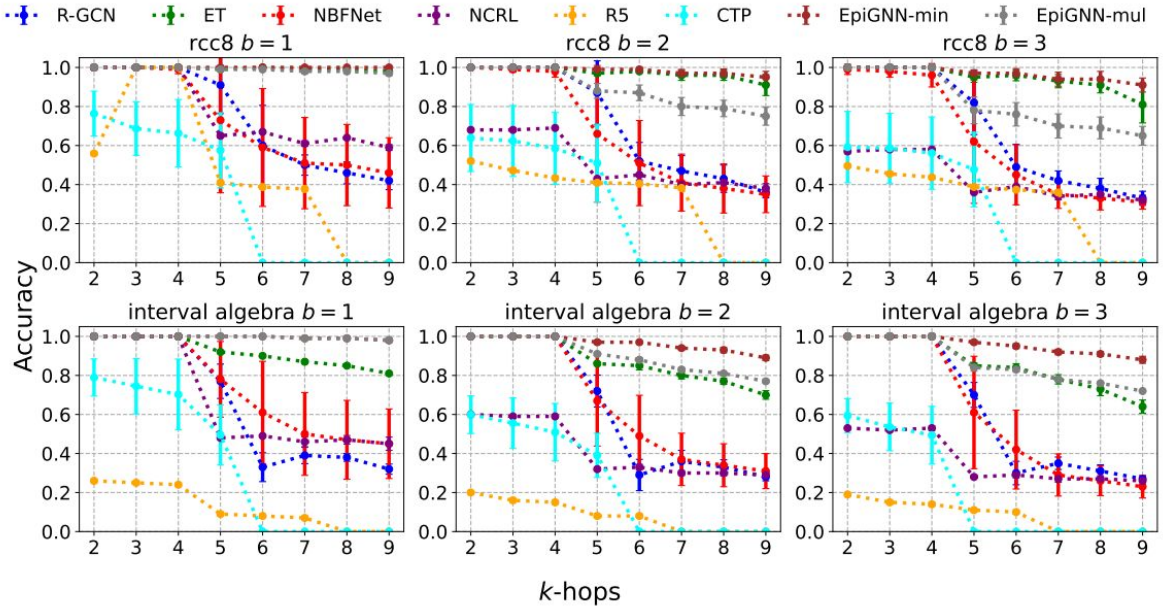
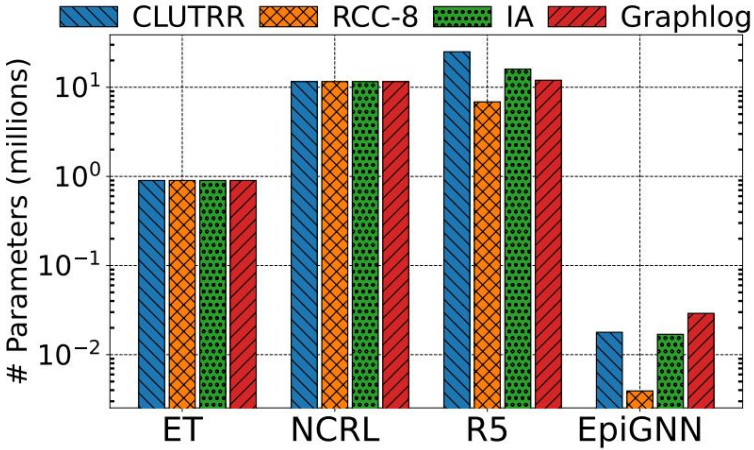


Figure 4: RCC-8 and Interval Algebra benchmark results (accuracy). R5 and CTP results for 5+ hops were set to zero since the model took longer than 30 minutes for inference. Models are trained on graphs with $b \in \{1, 2, 3\}$ paths of length $k \in \{2, 3, 4\}$. The best model for all cases is EpiGNN-`min`.

Ablations highlight the necessity of all the model's architectural propositions and EpiGNN is 100x parameter-efficient wrt. baselines

	CLUTRR		RCC-8	
	Avg	**Hard**	**Avg**	**Hard**
EpiGNN	0.99	0.99	0.96	0.80
- With facets=1	0.94	0.85	0.92	0.68
- Unconstrained embeddings	0.36	0.30	0.38	0.21
- MLP+distmul composition	0.29	0.31	0.13	0.13
- Forward model only	0.94	0.82	0.84	0.51



Outlook

Takeaways

1. Systematic reasoning enables generalization beyond training data. Could unlock high quality low-data learning.
2. NNs are bad at it because they lack the inductive biases that align their architecture with an algorithm that solves it.
3. For multi-path disjunctive reasoning (generalising single-path), we can align a relational GNN with the algebraic closure algorithm to yield SoTA results.

Thank you!

code: https://github.com/erg0dic/gnn-sg
STaR dataset: https://huggingface.co/datasets/erg0dic/STaR