How to Turn Your Knowledge Graph Embeddings into Generative Models

Lorenzo Loconte
University of Edinburgh, UK
loreloc.github.io

Nicola Di Mauro
University of Bari, Italy

Robert Peharz
TU Graz, Austria

Antonio Vergari
University of Edinburgh, UK

9th July 2025

Accepted to NeurIPS 2023 (as oral!)

Outline

1 ***A family of generative models of triples in KGs***

based on KGEs and probabilistic circuits

2 ***Reliable link prediction with logical constraints***

ensuring trustworthiness w.r.t. background knowledge

3 ***Efficiency and experimental results***

train for hours rather than days, yet be competitive

Outline

1 ***A family of generative models of triples in KGs***

based on KGEs and probabilistic circuits

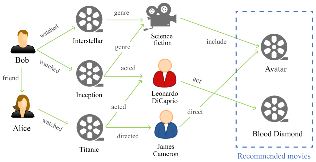
2 ***Reliable link prediction with logical constraints***

ensuring trustworthiness w.r.t. background knowledge

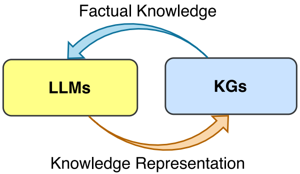
3 ***Efficiency and experimental results***

train for hours rather than days, yet be competitive

Knowledge graphs



Item recommendation



Augment LLMs

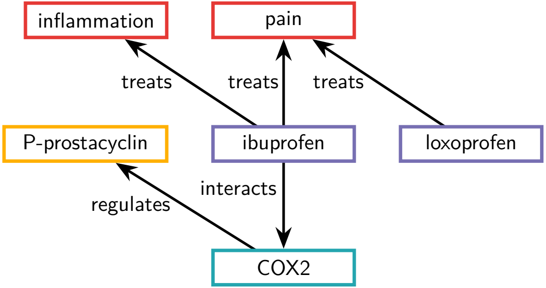
ChEMBL HGNC GOSTAR Ensembl BioGrid EBI Expression Atlas DisGeNET OMIM

Drug discovery

Guo et al., "A Survey on Knowledge Graph-Based Recommender Systems", 2020

Pan et al., "Unifying Large Language Models and Knowledge Graphs: A Roadmap", 2023

Gogleva et al., "Knowledge Graph-based Recommendation Framework Identifies [...] Resistance in [...] Cell Lung Cancer", 2021



- Drugs
- Symptoms
- Proteins
- Functions

$\langle$loxoprofen, treats, pain$\rangle$

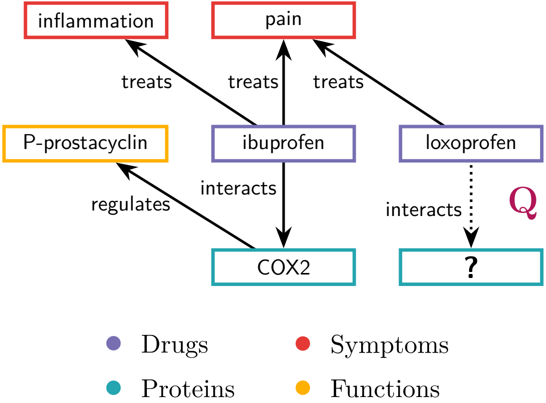
$\langle$ibuprofen, treats, pain$\rangle$

$\vdots$

$\langle$COX2, regulates, P-prostacyclin$\rangle$

$\langle$ibuprofen, interacts, COX2$\rangle$

Walsh, Mohamed, and Novácek, “BioKG: A Knowledge Graph for Relational Learning On Biological Data”, 2020



$\langle$loxoprofen, treats, pain$\rangle$

$\langle$ibuprofen, treats, pain$\rangle$

$\vdots$

$\langle$COX2, regulates, P-prostacyclin$\rangle$

$\langle$ibuprofen, interacts, COX2$\rangle$

$\mathcal{Q}$: $\langle$loxoprofen, interacts, **?**$\rangle$

Walsh, Mohamed, and Novácek, "BioKG: A Knowledge Graph for Relational Learning On Biological Data", 2020

KGE models

Knowledge graph embeddings (KGE) models such as ...

Complex Embeddings for Simple Link Prediction

Théo Trouillon, Johannes Welbl, +2 authors Guillaume Bouchard • Published in International Conference on... 19 June 2016 •

2,142 Citations

Highly Influential Citations 576

KGE models

Knowledge graph embeddings (KGE) models such as …

Complex Embeddings for Simple Link Prediction

Théo Trouillon, Johannes Welbl, +2 authors Guillaume Bouchard • Published in International Conference on... 19 June 2016 •

$$\phi_{\text{ComplEx}}(s, r, o) := f(\mathbf{e}_s, \mathbf{w}_r, \mathbf{e}_o) \in \mathbb{R} \qquad \mathbf{e}_s, \mathbf{w}_r, \mathbf{e}_o \in \mathbb{C}^d$$

KGE models

Knowledge graph embeddings (KGE) models such as ...

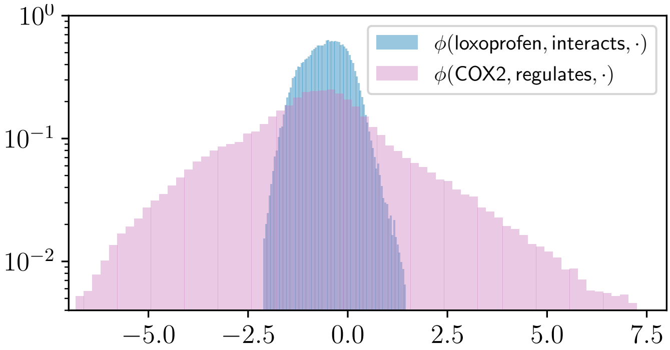
Complex Embeddings for Simple Link Prediction

Théo Trouillon, Johannes Welbl, +2 authors Guillaume Bouchard • Published in International Conference on... 19 June 2016 •

$$\phi_{\text{ComplEx}}(s, r, o) := f(\mathbf{e}_s, \mathbf{w}_r, \mathbf{e}_o) \in \mathbb{R} \qquad \mathbf{e}_s, \mathbf{w}_r, \mathbf{e}_o \in \mathbb{C}^d$$

$$\begin{array}{lll} 1^{\text{st}} & \phi_{\text{ComplEx}}(\text{loxoprofen}, \text{interacts}, \textbf{phosp-acid}) = & \mathbf{1.17} \Longleftarrow \\ 2^{\text{nd}} & \phi_{\text{ComplEx}}(\text{loxoprofen}, \text{interacts}, \text{COX2}) = & 0.95 \\ \vdots & & \end{array}$$

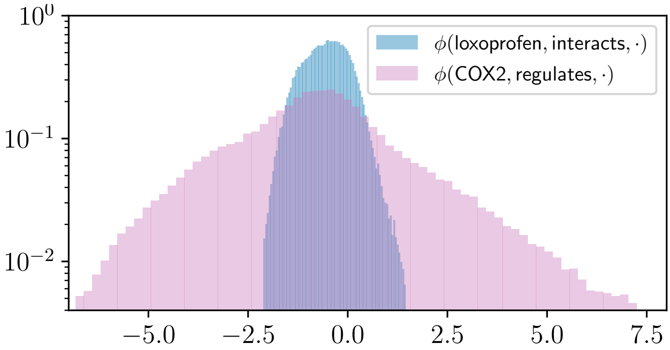
"Real-valued scores are difficult to interpret and compare"



Scores have different orders of magnitude...

Arakelyan, Minervini, and Augenstein, Adapting Neural Link Predictors for Complex Query Answering, 2023

Zhu et al., “A Closer Look at Probability Calibration of Knowledge Graph Embedding”, 2023



We would like *triples probabilities* instead !

Arakelyan, Minervini, and Augenstein, Adapting Neural Link Predictors for Complex Query Answering, 2023

Zhu et al., "A Closer Look at Probability Calibration of Knowledge Graph Embedding", 2023

Solutions! (1/3)

I ***Generative models of triples*** **(GeKCs)**

calibrated probabilistic predictions by modelling $p(S, R, O)$

sampling of new triples (more later!)

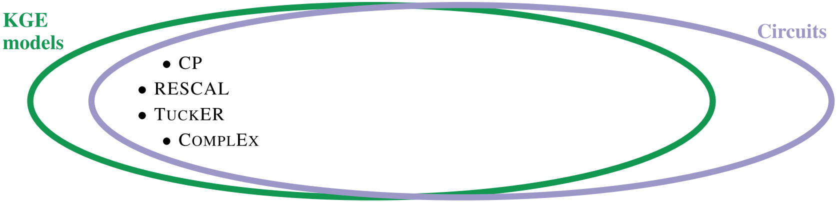
From KGE models ...

KGE models

- CP
- RESCAL
- TuckER
- ComplEx

Lacroix, Usunier, and Obozinski, "Canonical Tensor Decomposition for Knowledge Base Completion", 2018
Nickel, Tresp, and Kriegel, "A Three-Way Model for Collective Learning on Multi-Relational Data", 2011
Balazevic, Allen, and Hospedales, "TuckER: Tensor Factorization for Knowledge Graph Completion", 2019

From KGE models to circuits ...



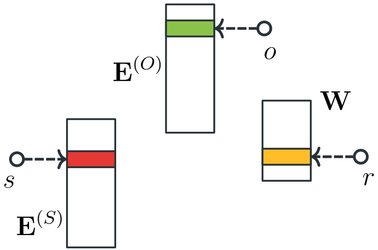
Lacroix, Usunier, and Obozinski, “Canonical Tensor Decomposition for Knowledge Base Completion”, 2018

Nickel, Tresp, and Kriegel, “A Three-Way Model for Collective Learning on Multi-Relational Data”, 2011

Balazevic, Allen, and Hospedales, “TuckER: Tensor Factorization for Knowledge Graph Completion”, 2019

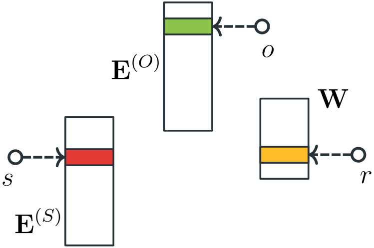
Canonical Polyadic (CP) KGE as a circuit

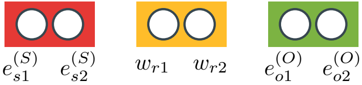
$$\phi(s, r, o) = \sum_{i=1}^{R} e_{si}^{(S)} \, w_{ri} \, e_{oi}^{(O)}$$



Canonical Polyadic (CP) KGE as a circuit

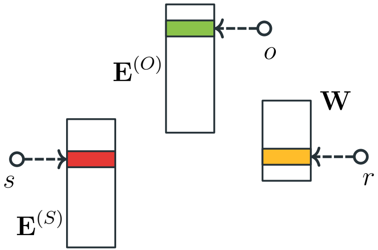
$$\phi(s, r, o) = \sum_{i=1}^{R} e^{(S)}_{si} \; w_{ri} \; e^{(O)}_{oi}$$

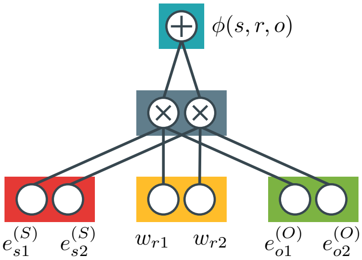




Canonical Polyadic (CP) KGE as a circuit

$$\phi(s, r, o) = \sum_{i=1}^{R} e_{si}^{(S)} \, w_{ri} \, e_{oi}^{(O)}$$



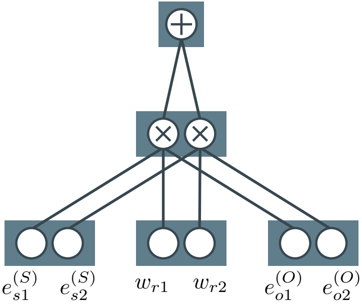


A circuit scoring triples

$$\mathbf{E}^{(S)} = \begin{bmatrix} 0.1 & 1.2 \\ 3.5 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 2.5 & 0.0 \\ -3.4 & -0.5 \\ -0.1 & 2.2 \end{bmatrix}$$

$$\mathbf{E}^{(O)} = \begin{bmatrix} -2.3 & 1.0 \\ 0.8 & -2.4 \\ 0.7 & 1.5 \end{bmatrix}$$



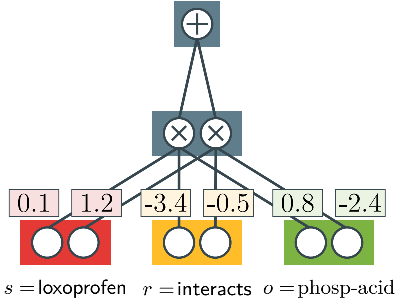
$$\phi_{\mathsf{CP}}(s,r,o) = \sum_{i=1}^{R} e^{(S)}_{si}\, w_{ri}\, e^{(O)}_{oi}$$

A circuit scoring triples

$$\mathbf{E}^{(S)} = \begin{bmatrix} 0.1 & 1.2 \\ 3.5 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 2.5 & 0.0 \\ -3.4 & -0.5 \\ -0.1 & 2.2 \end{bmatrix}$$

$$\mathbf{E}^{(O)} = \begin{bmatrix} -2.3 & 1.0 \\ 0.8 & -2.4 \\ 0.7 & 1.5 \end{bmatrix}$$



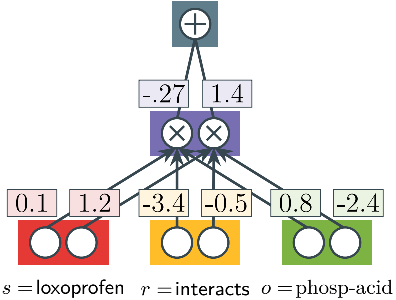
$$\phi_{\text{CP}}(\text{loxoprofen, interacts, phosp-acid}) = \sum_{i=1}^{R} e^{(S)}_{\text{loxoprofen},i} \, w_{\text{interacts},i} \, e^{(O)}_{\text{phosp-acid},i}$$

A circuit scoring triples

$$\mathbf{E}^{(S)} = \begin{bmatrix} \boxed{0.1} & \boxed{1.2} \\ 3.5 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 2.5 & 0.0 \\ \boxed{-3.4} & \boxed{-0.5} \\ -0.1 & 2.2 \end{bmatrix}$$

$$\mathbf{E}^{(O)} = \begin{bmatrix} -2.3 & 1.0 \\ \boxed{0.8} & \boxed{-2.4} \\ 0.7 & 1.5 \end{bmatrix}$$



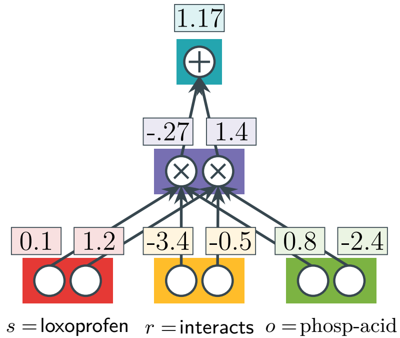
$$\phi_{\mathsf{CP}}(\text{loxoprofen}, \text{interacts}, \text{phosp-acid}) = \sum_{i=1}^{R} e^{(S)}_{\text{loxoprofen},i} \, w_{\text{interacts},i} \, e^{(O)}_{\text{phosp-acid},i}$$

A circuit scoring triples

$$\mathbf{E}^{(S)} = \begin{bmatrix} 0.1 & 1.2 \\ 3.5 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$$

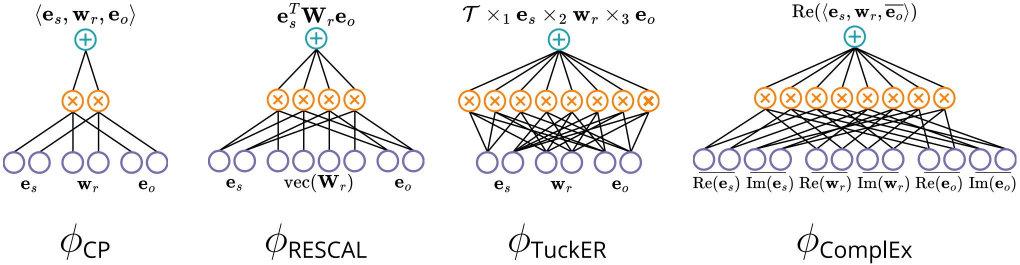
$$\mathbf{W} = \begin{bmatrix} 2.5 & 0.0 \\ -3.4 & -0.5 \\ -0.1 & 2.2 \end{bmatrix}$$

$$\mathbf{E}^{(O)} = \begin{bmatrix} -2.3 & 1.0 \\ 0.8 & -2.4 \\ 0.7 & 1.5 \end{bmatrix}$$

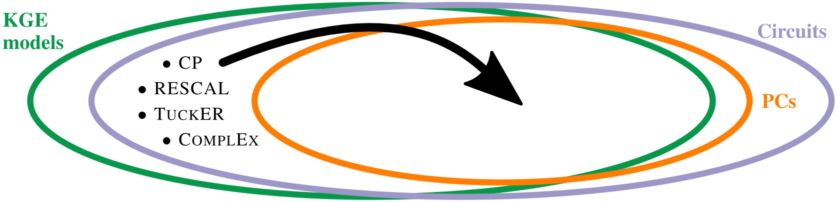


$$\phi_{\text{CP}}(\text{loxoprofen}, \text{interacts}, \text{phosp-acid}) = \sum_{i=1}^{R} e^{(S)}_{\text{loxoprofen},i} \; w_{\text{interacts},i} \; e^{(O)}_{\text{phosp-acid},i}$$

From KGE models to circuits ...

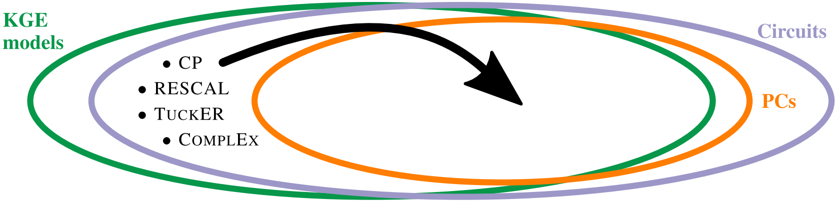


... to probabilistic circuits



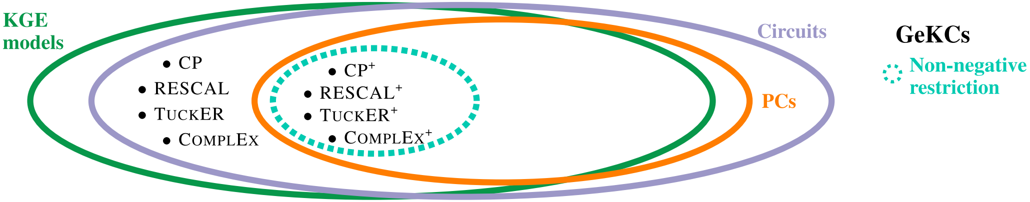
From scores $\phi(s, r, o)$ to *triple probabilities* $p(s, r, o)$

... to probabilistic circuits



1. Ensure $\phi(s, r, o) \geq 0$, $\quad p(s, r, o) = \frac{1}{Z} \cdot \phi(s, r, o)$

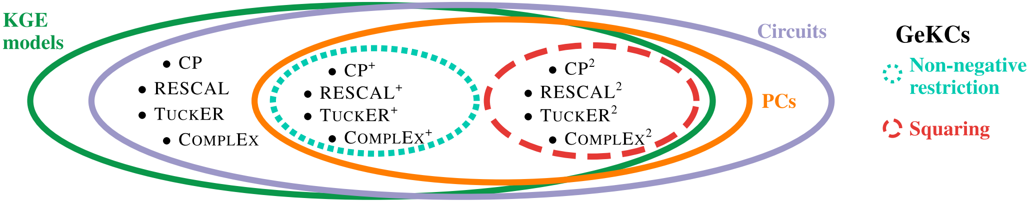
… to probabilistic circuits



Enforce *non-negative embeddings*

$\Longrightarrow$ **Less accurate on link prediction …**

… *to probabilistic circuits*



Square score functions (unrestricted embeddings)

$\Longrightarrow$ **Competitive on link prediction !**

Squared circuits

$$p(\mathbf{x}) = \frac{1}{Z}\, \phi^2(\mathbf{x}) = \frac{1}{Z}\, \phi(\mathbf{x}) \cdot \phi(\mathbf{x}), \qquad \phi(\mathbf{x}) \in \mathbb{R}$$

where parameters and input functions can be **negative**

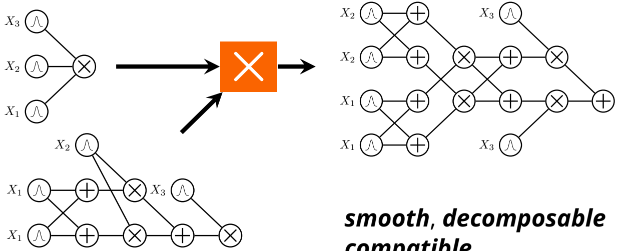
Loconte et al., “Subtractive Mixture Models via Squaring: Representation and Learning”, 2024

Loconte, Mengel, and Vergari, “Sum of Squares Circuits”, 2025

Squared circuits

$$p(\mathbf{x}) = \frac{1}{Z}\,\phi^2(\mathbf{x}) = \frac{1}{Z}\,\phi(\mathbf{x}) \cdot \phi(\mathbf{x}), \qquad \phi(\mathbf{x}) \in \mathbb{R}$$

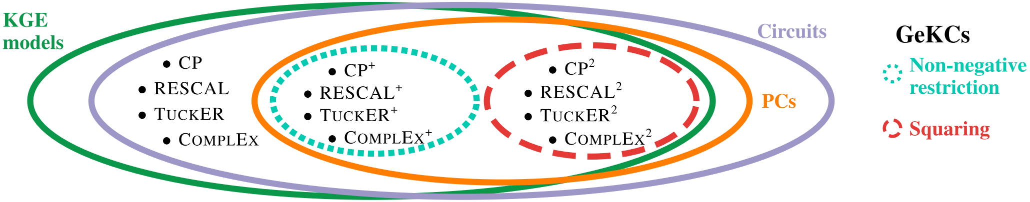
Tractable product



smooth**, **decomposable
compatible

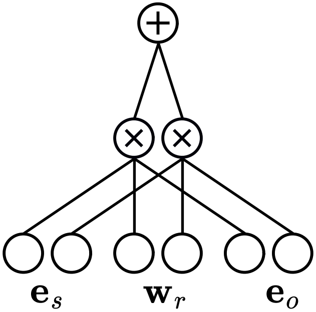
Vergari et al., “A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference”, 2021

... to probabilistic circuits



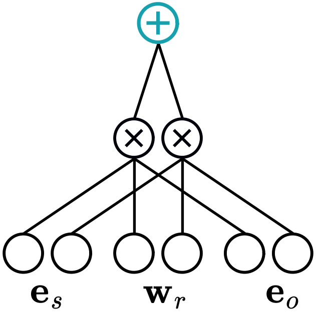
1. **Ensure** $\phi(s,r,o) \geq 0$, $\quad p(s,r,o) = \frac{1}{Z} \cdot \phi(s,r,o)$

2. **Computation of** $Z = \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \phi(s,r,o)$



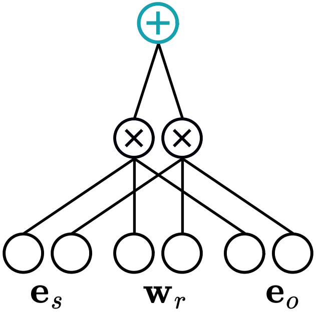
$$Z = \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \phi_{\mathsf{CP}^+}(s, r, o)$$

The summation over triples computing Z ...



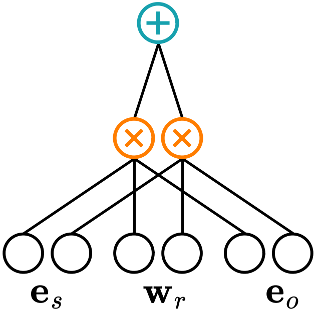
$$Z = \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \sum_{i=1}^{d} e_{si}^{(S)} w_{ri} e_{oi}^{(O)}$$

The summation over triples computing Z …



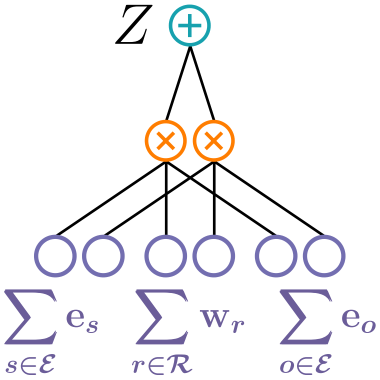
$$Z = \sum_{i=1}^{d} \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} e_{si}^{(S)} w_{ri} e_{oi}^{(O)}$$

... can be pushed (smoothness) ....



$$Z = \sum_{i=1}^{d} \left(\sum_{s \in \mathcal{E}} e_{si}^{(S)} \right) \times \left(\sum_{r \in \mathcal{R}} w_{ri} \right) \times \left(\sum_{o \in \mathcal{E}} e_{oi}^{(O)} \right)$$

… and broken down (decomposability) …



... thus requiring linear time w.r.t. $|\mathcal{E}|$

Outline

1 ***A family of generative models of triples in KGs***

based on KGEs and probabilistic circuits

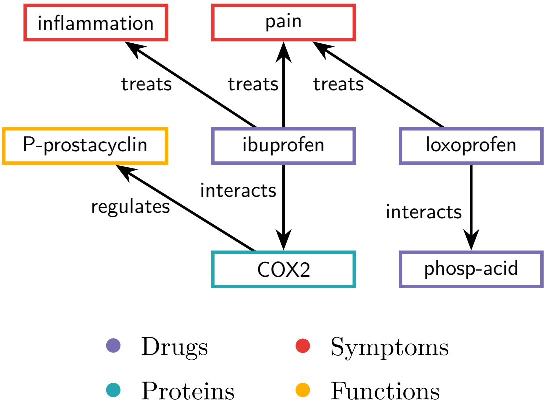
2 ***Reliable link prediction with logical constraints***

ensuring trustworthiness w.r.t. background knowledge

3 ***Efficiency and experimental results***

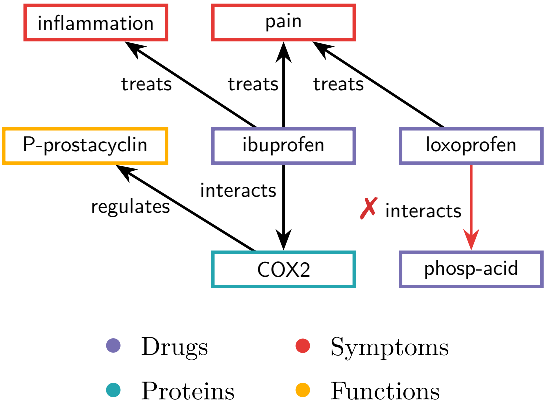
train for hours rather than days, yet be competitive

“KGE models predictions violate even simple logical constraints”



$\mathbf{Q}:\ \langle \text{loxoprofen, interacts, } \mathbf{?} \rangle$

$\mathbf{A}:\ \langle \text{loxoprofen, interacts, } \textbf{phosp-acid} \rangle$

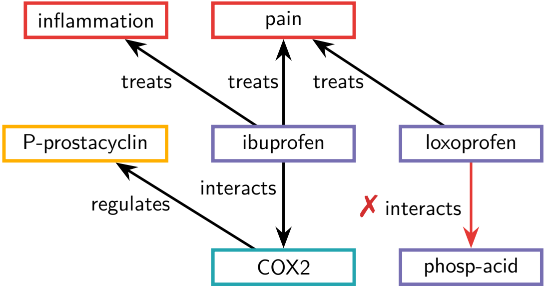


$\mathbf{Q}:$ ⟨loxoprofen, interacts, **?**⟩

$\mathbf{A}:$ ⟨loxoprofen, interacts, **phosp-acid**⟩

✗

“ *interacts* ” can only hold between ***drugs*** and ***proteins***



- Drugs
- Proteins
- Symptoms
- Functions

Q : ⟨loxoprofen, interacts, **?**⟩

A : ⟨loxoprofen, interacts, **phosp-acid**⟩

✗

" *interacts* " can only hold between ***drugs*** and ***proteins***

ComplEx predicts a triple violating a constraint !

Solutions! (2/3)

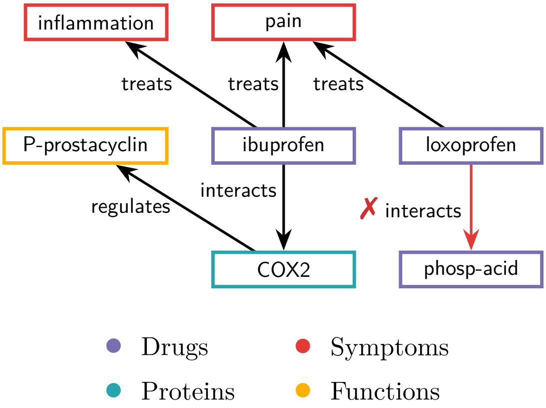
I ***Generative models for KGs*** **(GeKCs)**

calibrated probabilistic predictions by modelling $p(S, R, O)$

sampling of new triples (more later!)

II ***Integrate constraints with guarantees***

such as the domain schema



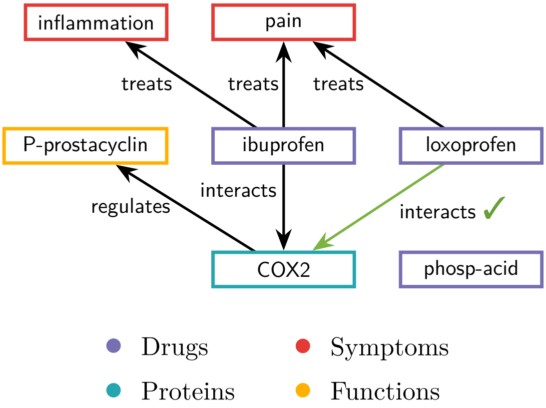
Q: $\langle$loxoprofen, interacts, **?**$\rangle$

A: $\langle$loxoprofen, interacts, **phosp-acid**$\rangle$

✗

"*interacts*" can only hold between ***drugs*** and ***proteins***

$p($**loxoprofen**, interacts, **phosp-acid**$) = 0$



$\mathbf{Q}:$ $\langle$loxoprofen, interacts, **?**$\rangle$

$\mathbf{A}:$ $\langle$loxoprofen, interacts, **COX2**$\rangle$

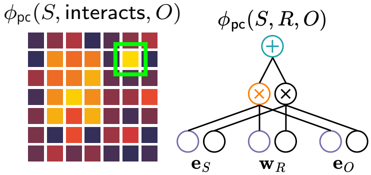
✓

“ *interacts* ” can only hold between ***drugs*** and ***proteins***

$p($**loxoprofen**, interacts, **phosp-acid**$) = 0$

$p($**loxoprofen**, interacts, **COX2**$) > 0$

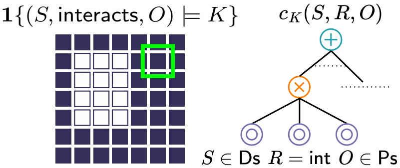
GeKCs for logical constraints

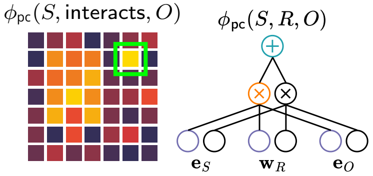


$$p_K\left(\square\ \mathsf{loxoprofen}, \mathsf{interacts}, \mathbf{phosp\text{-}acid}\right) = 0$$

Ahmed et al., "Semantic probabilistic layers for neuro-symbolic learning", 2022

GeKCs for logical constraints

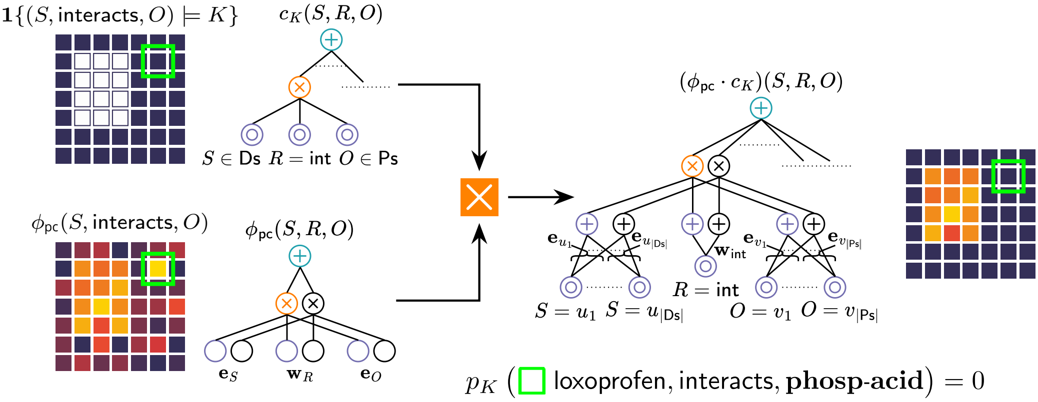




$p_K\left(\text{loxoprofen}, \text{interacts}, \textbf{phosp-acid}\right) = 0$

Ahmed et al., "Semantic probabilistic layers for neuro-symbolic learning", 2022

GeKCs for logical constraints



Outline

1 ***A family of generative models of triples in KGs***

based on KGEs and probabilistic circuits

2 ***Reliable link prediction with logical constraints***

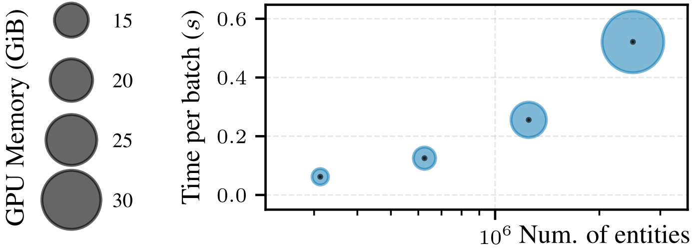
ensuring trustworthiness w.r.t. background knowledge

3 ***Efficiency and experimental results***

train for hours rather than days, yet be competitive

“Training on relatively large knowledge graphs is expensive”

Some benchmarks...



WIKIDATA

$107 \cdot 10^6$ entities

Vrandecić and Krötzsch, “Wikidata: a free collaborative knowledgebase”, 2014

Solutions! (3/3)

I

Generative models for KGs **(GeKCs)**

calibrated probabilistic predictions by modelling $p(S, R, O)$

sampling of new triples (more later!)

II

Integrate constraints with guarantees

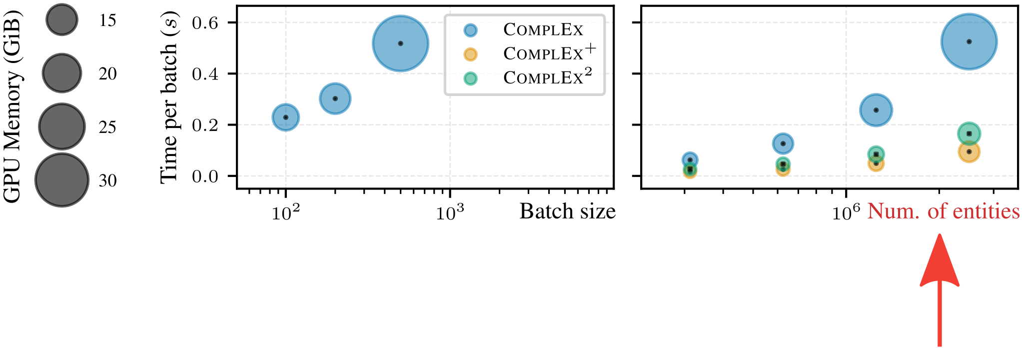
such as the domain schema

III

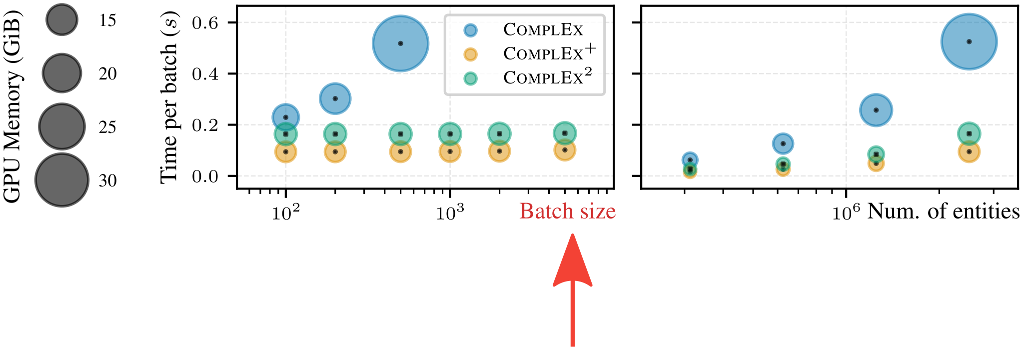
Scale to KGs with millions of entities and triples

using probabilistic training objectives

Speed-up training on large KGs



Speed-up training on large KGs



Learning ...

... by discriminative objectives (*pseudo-log-likelihood*)

$$\mathcal{L}_{\mathsf{PLL}} := \sum_{(s,r,o)\in\mathcal{D}} w_s \log p(s \mid r, o) + w_r \log p(r \mid s, o) + w_o \log p(o \mid s, r)$$

Ruffinelli, Broscheit, and Gemulla, "You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings", 2020
Chen et al., "Relation Prediction as an Auxiliary Training Objective for Improving Multi-Relational Graph Representations", 2021

Learning ...

... by discriminative objectives (*pseudo-log-likelihood*)

$$\mathcal{L}_{\mathsf{PLL}} := \sum_{(s,r,o)\in\mathcal{D}} w_s \log p(s \mid r, o) + w_r \log p(r \mid s, o) + w_o \log p(o \mid s, r)$$

... by *maximum-log-likelihood* estimation

$$\mathcal{L}_{\mathsf{MLE}} := \sum_{(s,r,o)\in\mathcal{D}} \log p(s, r, o) = -|\mathcal{D}| \log Z + \sum_{(s,r,o)\in\mathcal{D}} \log \phi_{\mathsf{pc}}(s, r, o)$$

(faster, as marginalization require a single circuit evaluation)

Ruffinelli, Broscheit, and Gemulla, “You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings”, 2020

Chen et al., “Relation Prediction as an Auxiliary Training Objective for Improving Multi-Relational Graph Representations”, 2021

Mean Reciprocal Rank (MRR) $\uparrow$

Model	FB15k-237	WN18RR	ogbl-biokg
CP	0.310	**0.105**	0.831
CP^+	0.237	0.027	0.496
CP^2	**0.315**	**0.104**	**0.848**
ComplEx	**0.342**	**0.471**	0.829
ComplEx^+	0.214	0.030	0.503
ComplEx^2	0.334	0.420	**0.858**

Mean Reciprocal Rank (MRR) ↑

Model	FB15k-237	WN18RR	ogbl-biokg
CP	0.310	**0.105**	0.831
CP^{+}	0.237	0.027	0.496
CP^{2}	**0.315**	**0.104**	**0.848**
ComplEx	**0.342**	**0.471**	0.829
$ComplEx^{+}$	0.214	0.030	0.503
$ComplEx^{2}$	0.334	0.420	**0.858**

GeKCs are competitive with KGE models ...

Mean Reciprocal Rank (MRR) ↑

Model	FB15k-237	WN18RR	ogbl-biokg
CP	0.310	**0.105**	0.831
CP^+	0.237	0.027	0.496
CP^2	**0.315**	**0.104**	**0.848**
ComplEx	**0.342**	**0.471**	0.829
$ComplEx^+$	0.214	0.030	0.503
$ComplEx^2$	0.334	0.420	**0.858**

… *and achieve the best results on* ogbl-biokg

Sampling triples

Kernel triple distance **to measure their quality**

Binkowski et al., "Demystifying MMD GANs", 2018

Sampling triples

Kernel triple distance to measure their quality

Empirical KTD $\downarrow$

Model	FB15k-237		WN18RR		ogbl-biokg	
Uniform	0.589		0.766		1.822	
	PLL	MLE	PLL	MLE	PLL	MLE
ComplEx^2	0.326	**0.102**	0.338	**0.278**	0.104	**0.034**

Binkowski et al., “Demystifying MMD GANs”, 2018

Takeaways

I *A generative perspective of KGE models* (GeKCs)

II *Reliable predictions with logical constraints*

III *Speed-up training and reduce costs*

Takeaways

Questions?

I *A generative perspective of KGE models* (GeKCs)

II *Reliable predictions with logical constraints*

III *Speed-up training and reduce costs*

More link prediction benchmarks

Mean Reciprocal Rank (MRR) $\uparrow$

Model	FB15k-237		WN18RR		ogbl-biokg	
	PLL	MLE	PLL	MLE	PLL	MLE
CP	0.310	—	**0.105**	—	0.831	—
CP^+	0.237	0.230	0.027	0.026	0.496	0.501
CP^2	**0.315**	0.282	**0.104**	0.091	**0.848**	0.829
ComplEx	**0.342**	—	**0.471**	—	0.829	—
$ComplEx^+$	0.214	0.205	0.030	0.029	0.503	0.516
$ComplEx^2$	0.334	0.300	0.420	0.391	**0.858**	0.840

Instantiate GeKCs from KGE models

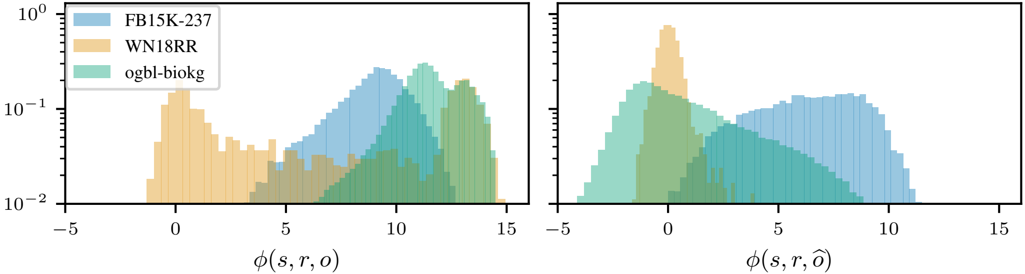
(as a way to initialize the parameters)

Mean Reciprocal Rank (MRR) ↑

Model	FB15k-237		WN18RR		ogbl-biokg	
	PLL	MLE	PLL	MLE	PLL	MLE
ComplEx	**0.344**	—	**0.470**	—	0.829	—
ComplEx^2	0.333	0.301	0.416	0.390	0.859	0.839
ComplEx^2 ⋆	**0.342**	**0.340**	**0.462**	**0.463**	0.859	0.828

Scores are mostly non-negative

(thus squaring them has little effect on the rankings)



Semantic consistency scores

Sem@1 scores ↑

Model	**Embedding size**			
	10	50	200	1000
ComplEx	99.68	99.90	99.93	**99.94**
ComplEx^2	82.50	94.22	99.30	99.50
D-ComplEx^2	**100.00**	**100.00**	**100.00**	**100.00**

Hubert et al., "New Strategies for Learning Knowledge Graph Embeddings: The Recommendation Case", 2022