How to Turn Your Knowledge Graph Embeddings into Generative Models

Lorenzo Loconte
University of Edinburgh, UK

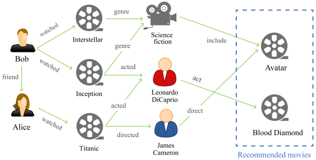
Nicola Di Mauro
University of Bari, Italy

Robert Peharz
TU Graz, Austria

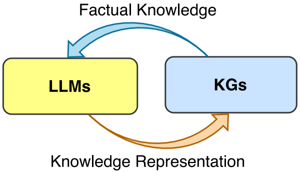
Antonio Vergari
University of Edinburgh, UK

NeurIPS 2023 Oral

Knowledge graphs



Item recommendation



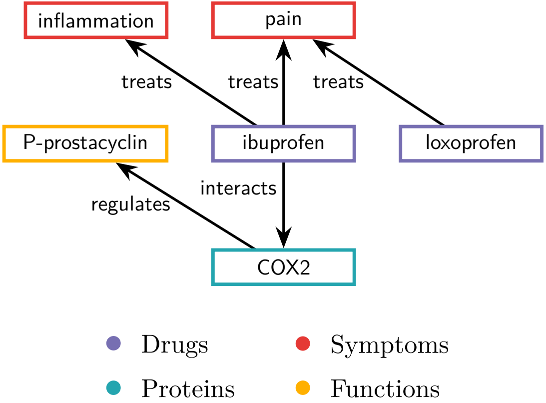
Augment LLMs

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



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

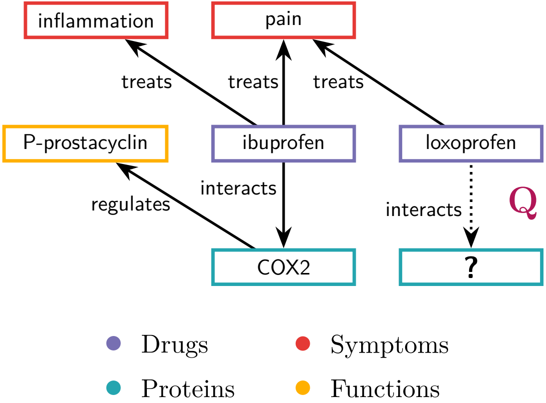
$\langle$ibuprofen, treats, pain$\rangle$

$\vdots$

$\langle$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$\rangle$

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 •

2,142 Citations

Highly Influential Citations 576

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

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 •

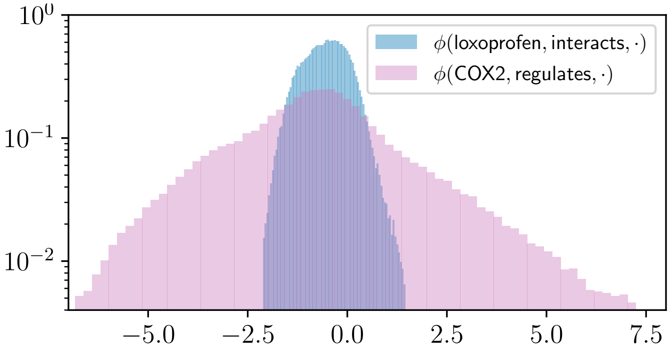
2,142 Citations

Highly Influential Citations 576

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

$$\begin{array}{llll} 1^{\text{st}} & \phi_{\text{ComplEx}}(\text{loxoprofen}, \text{interacts}, \textbf{phosp-acid}) & = & \mathbf{2.3} \Longleftarrow \\ 2^{\text{nd}} & \phi_{\text{ComplEx}}(\text{loxoprofen}, \text{interacts}, \text{COX2}) & = & 1.3 \\ \vdots & & & \end{array}$$

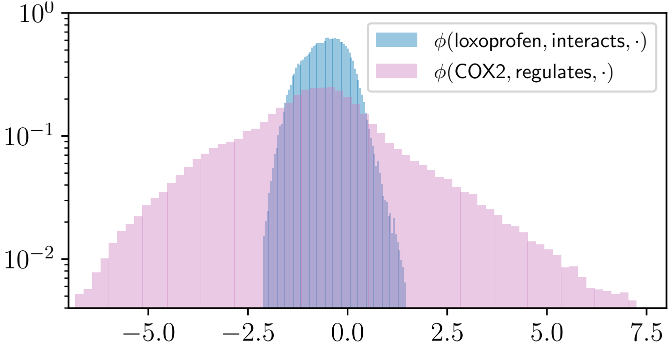
Scores …



… are difficult to interpret and compare

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

Scores ...



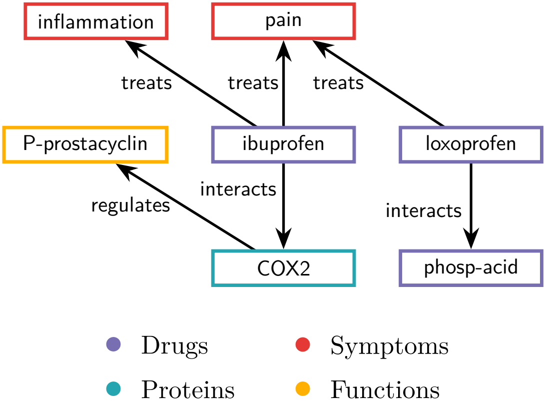
We would like *probabilities* instead !

Issues?

I

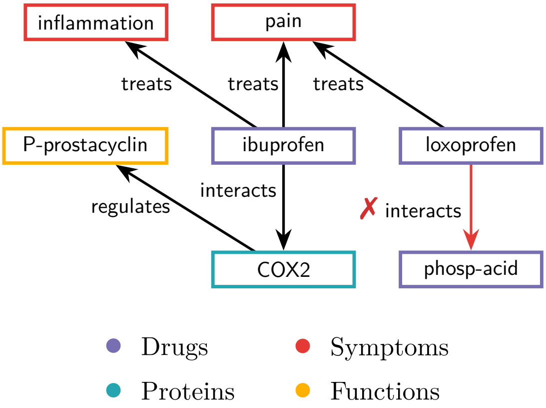
How to measure the confidence of predictions ?

and compare / combine scores



Q: ⟨loxoprofen, interacts, **?**⟩

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



Q : ⟨loxoprofen, interacts, **?**⟩

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

✗

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

Issues?

I

How to measure the confidence of predictions ?

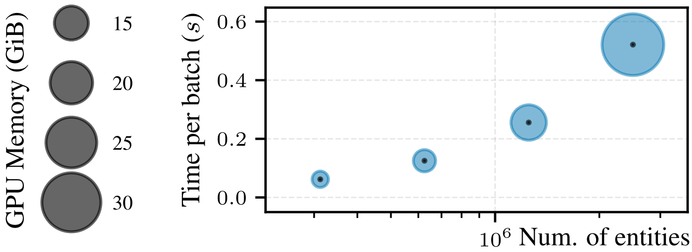
and compare / combine scores

II

How to guarantee the satisfaction of constraints ?

such as domain knowledge

Training is expensive



WIKIDATA

$107 \cdot 10^6$ entities

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

Issues?

I ***How to measure the confidence of predictions ?***

and compare / combine scores

II ***How to guarantee the satisfaction of constraints ?***

such as domain knowledge

III ***How to scale to KGs with millions of entities ?***

and be memory efficient

Solutions!

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

calibrated probabilistic predictions, sampling of new triples

From KGE models ...

KGE models

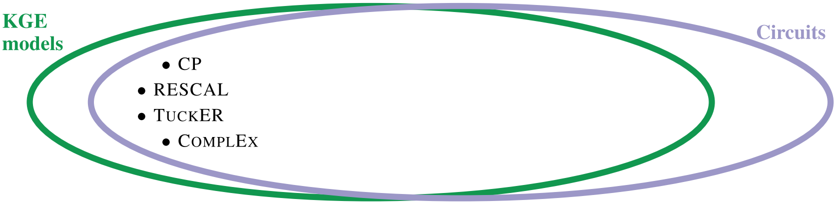
- CP
- RESCAL
- TuckER
- ComplEx

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

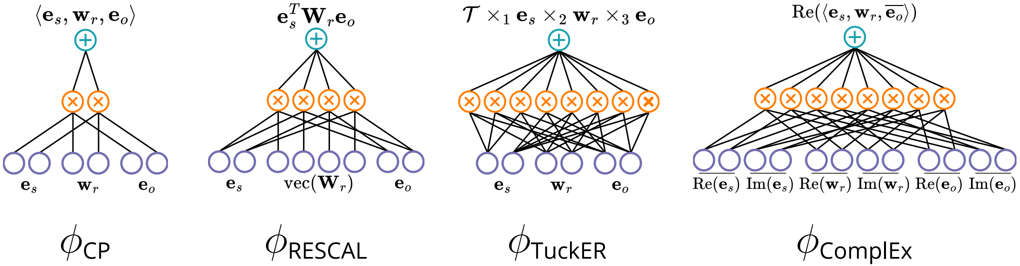
Balazevic, Allen, and Hospedales, “TuckER: Tensor Factorization for Knowledge Graph Completion”, 2019

From KGE models to circuits ...

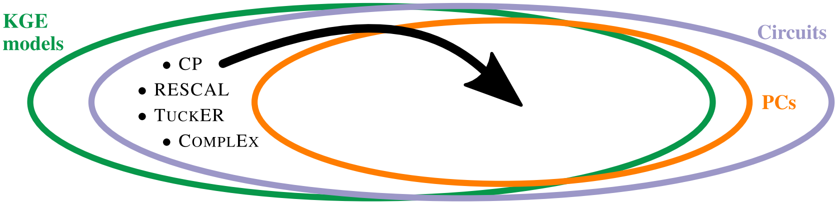


Lacroix, Usunier, and Obozinski, “Canonical Tensor Decomposition for Knowledge Base Completion”, 2018
Nickel, Tresp, and Kriegel, “A Three-Way Model for Collective Learning on Multi-Relational Data”, 2011
Balazevic, Allen, and Hospedales, “TuckER: Tensor Factorization for Knowledge Graph Completion”,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

From KGE models to circuits ...



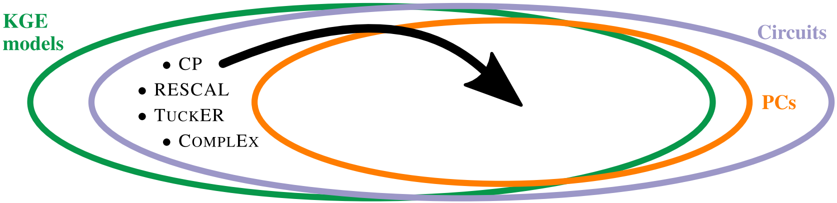
… to probabilistic circuits



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

Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020

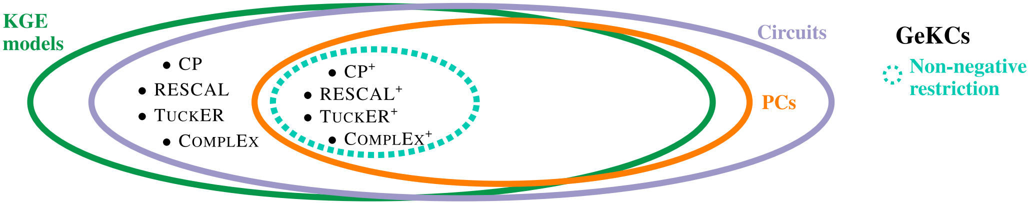
… to probabilistic circuits



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

Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020

... to probabilistic circuits

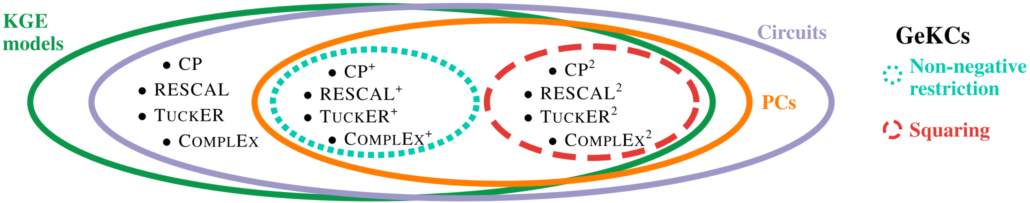


Enforce *non-negative embeddings*

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

Choi, Vergari, and Broeck, “Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling”, 2020

… to probabilistic circuits

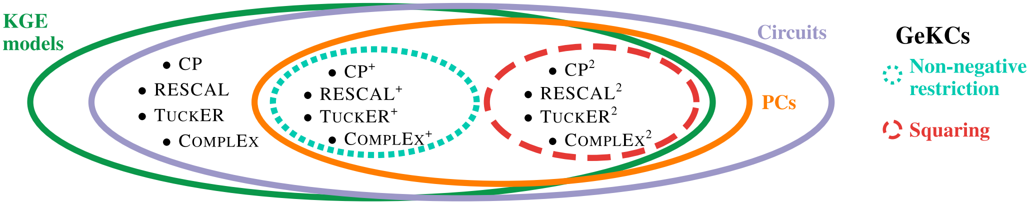


Square score functions (unrestricted embeddings)

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

Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020

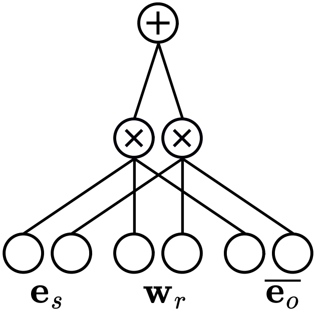
… to probabilistic circuits



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

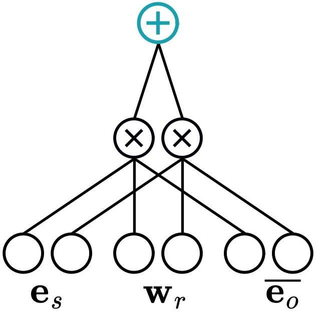
2. Computation of $Z = \sum_{s \in \mathcal{E},\, r \in \mathcal{R},\, o \in \mathcal{E}} \phi(s,r,o)$

Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020



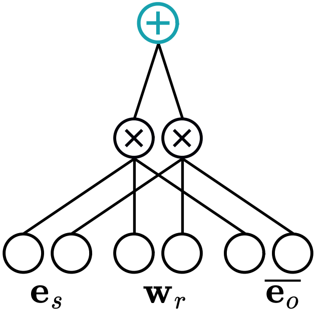
$$Z = \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \phi_{\text{ComplEx}^+}(s, r, o)$$

The summation over triples computing Z ...



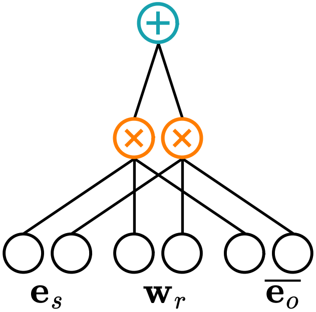
$$Z = \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \sum_{i=1}^{d} \mathrm{Re}\left(\mathbf{e}_{si}\mathbf{w}_{ri}\overline{\mathbf{e}_{oi}}\right)$$

The summation over triples computing Z ...



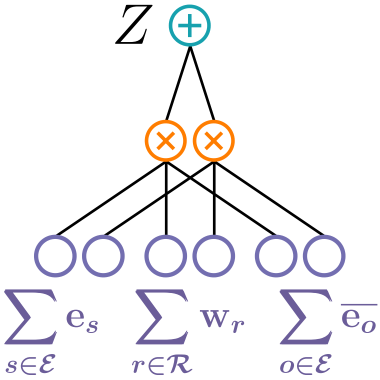
$$Z = \sum_{i=1}^{d} \sum_{s\in\mathcal{E},\, r\in\mathcal{R},\, o\in\mathcal{E}} \mathrm{Re}\left(\mathbf{e}_{si}\mathbf{w}_{ri}\overline{\mathbf{e}_{oi}}\right)$$

... can be pushed ....



$$Z = \sum_{i=1}^{d} \mathrm{Re}\left[\left(\sum_{s\in\mathcal{E}} \mathbf{e}_{si}\right) \times \left(\sum_{r\in\mathcal{R}} \mathbf{w}_{ri}\right) \times \left(\sum_{o\in\mathcal{E}} \overline{\mathbf{e}_{oi}}\right)\right]$$

… and broken down …



... thus requiring linear time !

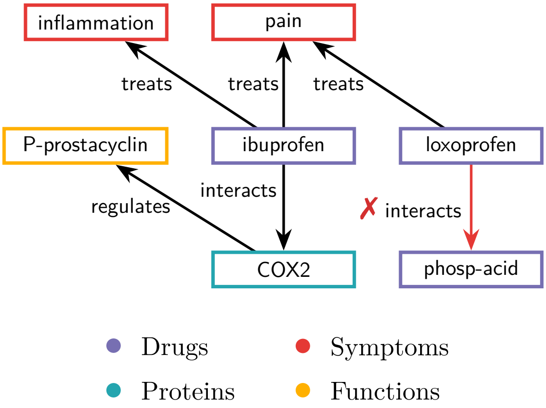
Solutions!

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

calibrated probabilistic predictions, sampling of new triples

II ***Integrate constraints with guarantees***

such as the domain schema



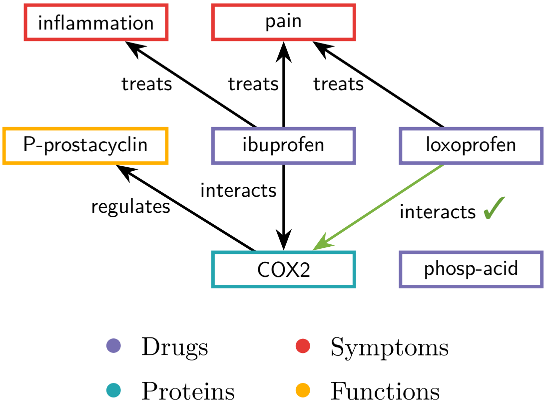
$\mathbf{Q}:$ ⟨loxoprofen, interacts, **?**⟩

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

✗

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

$p(\text{loxoprofen}, \text{interacts}, \text{phosp-acid}) = 0$



Q : ⟨loxoprofen, interacts, **?**⟩

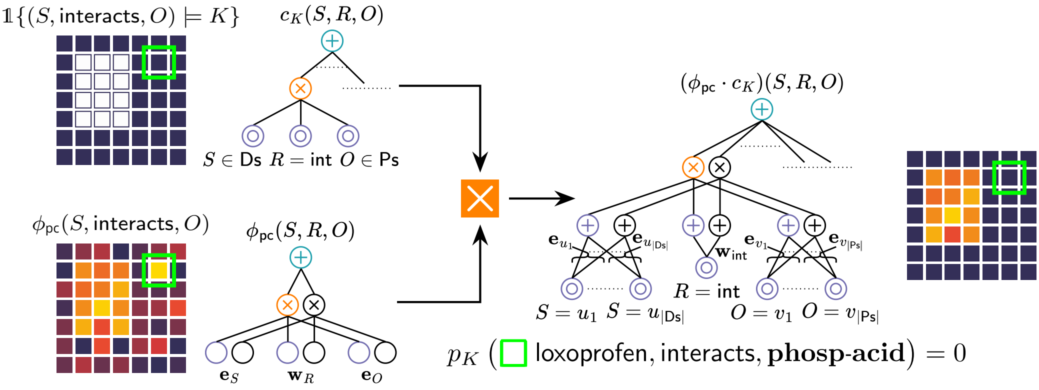
A : ⟨loxoprofen, interacts, **COX2**⟩

✓

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

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

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



Logical constraints + **GeKCs**

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

Solutions!

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

calibrated probabilistic predictions, sampling of new triples

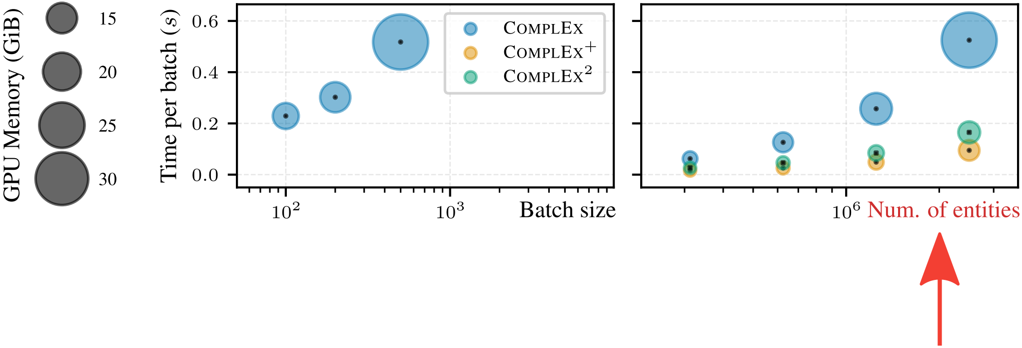
II ***Integrate constraints with guarantees***

such as the domain schema

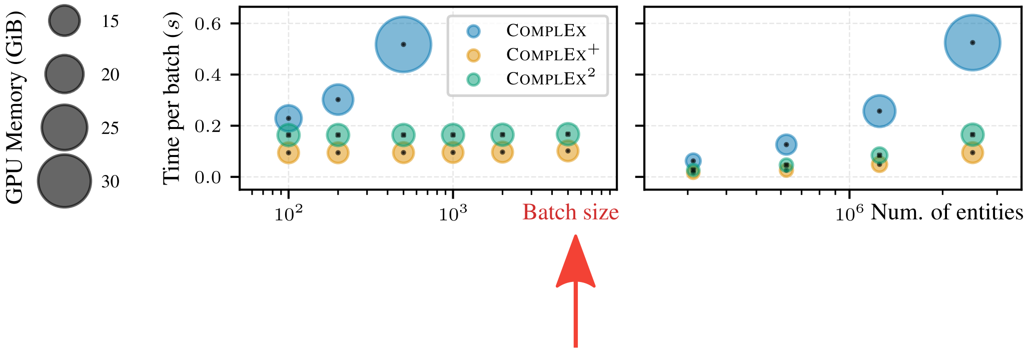
III ***Scale to KGs with millions of entities and triples***

speed-up training and save memory

Speed-up training on large KGs



Speed-up training on large KGs



Learning ...

... by discriminative objectives, generalised as a weighted *pseudo-log-likelihood*

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

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,
generalised as a weighted ***pseudo-log-likelihood***

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

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

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

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 ↓

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

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

II — *Reliable predictions with logical constraints*

III — *Speed-up training and reduce costs*

more on circuits

A. Vergari, Y. Choi, and R. Peharz
Probabilistic Circuits: representations, inference, learning and applications
Tutorial @ NeurIPS 2022

Z. Yu, M. Trapp and K. Kersting
Characteristic circuits
Oral @ NeurIPS 2023

Takeaways

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

II *Reliable predictions with logical constraints*

III *Speed-up training and reduce costs*

april

april-tools.github.io

about probabilities, reasoning, integrals & logic

Poster Session 1 **#1205**

Paper *Code*

@loreloc_

Link prediction benchmarks

Mean Reciprocal Rank (MRR) ↑

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

Mean Reciprocal Rank (MRR) $\uparrow$

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
$\text{ComplEx}^2\ \star$	**0.342**	**0.340**	**0.462**	**0.463**	0.859	0.828

Semantic consistency scores

Sem@k scores $\uparrow$

Model	k	**Embedding size** 10	50	200	1000
ComplEx	**1**	99.68	99.90	99.93	**99.94**
	20	99.81	99.79	99.85	99.91
	100	99.60	99.44	99.60	99.77
ComplEx^2	1	82.50	94.22	99.30	99.50
	20	86.50	96.70	99.42	99.64
	100	90.66	97.71	99.23	98.78
D-ComplEx^2	1	**100.00**	**100.00**	**100.00**	**100.00**
	20	**100.00**	**100.00**	**100.00**	**100.00**
	100	**100.00**	**100.00**	**100.00**	**100.00**

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

Logical constraints improve small GeKCs