How to Turn your Knowledge Graph Embeddings into Generative Models via Probabilistic Circuits

Lorenzo Loconte
University of Edinburgh, UK

Nicola Di Mauro
University of Bari, Italy

Robert Peharz
TU Graz, Austria

Antonio Vergari
University of Edinburgh, UK

TL;DR: *"We cast KGE models into generative models supporting exact and efficient marginalisation, sampling and the integration of propositional constraints with guarantees."*

From Knowledge Graph Embeddings ...

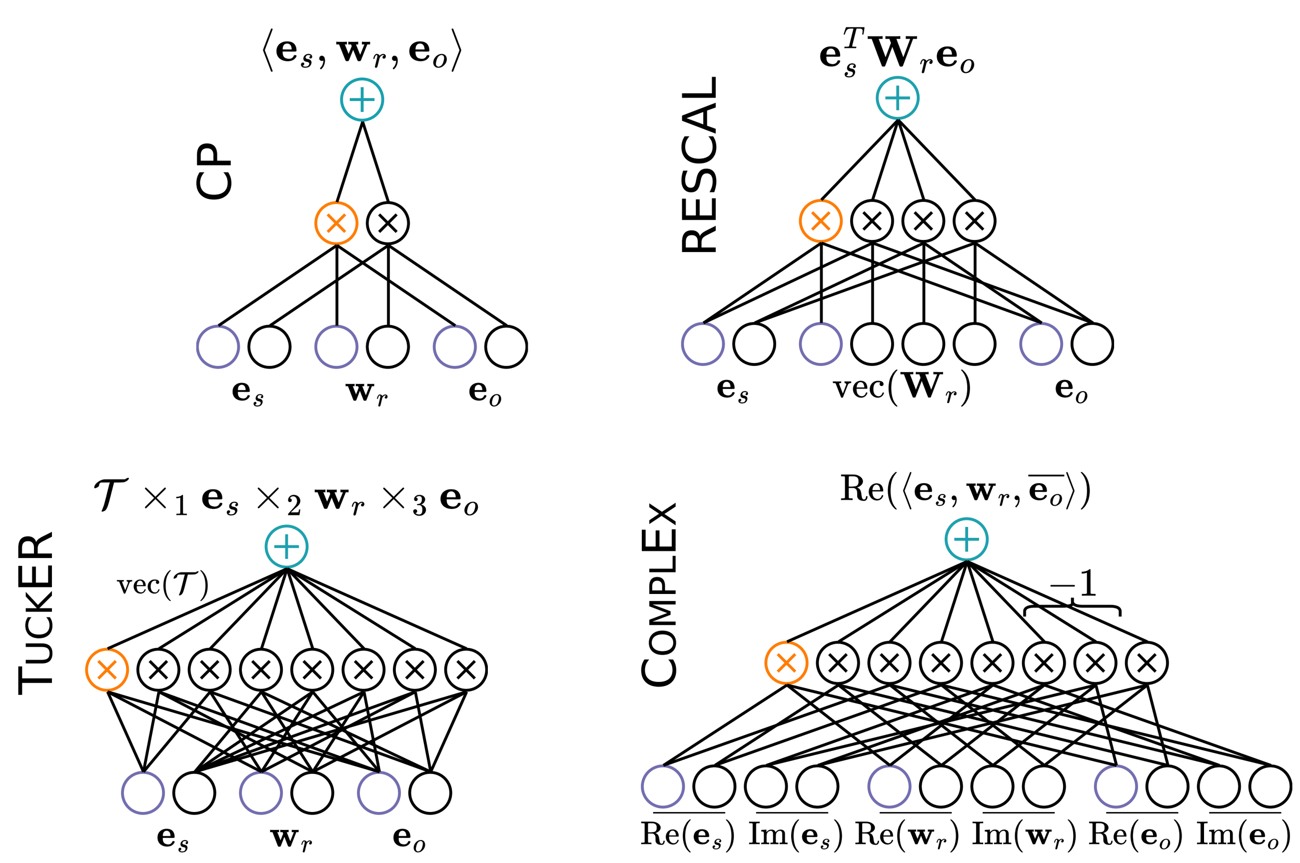
A Knowledge Graph Embeddings (KGE) model can be interpreted as an *energy-based model* defining an unnormalized distribution over triples (S, R, O)

$$p(S=s, R=r, O=o) = \frac{1}{Z}\exp\phi(s,r,o)$$

where $\phi : \mathcal{E}\times\mathcal{R}\times\mathcal{E}\to\mathbb{R}$ is a triple score function.

KGE models have several **shortcomings**:

1. Computing triple probabilities is impractical
   Z requires 10^{11} evaluations of ϕ on FB15K-237 [5]
   thus comparing triple scores is not easy
2. Learning by maximum-likelihood is infeasible
   requires contrastive objectives or
   expensive cross-entropies, e.g., $\log p(o \mid s, r)$ [4]
3. Logical constraints are violated at test time [3]
   e.g., $Q:(\text{loxoprofen}, \text{interacts}, ?)$
   $A:\text{phosphoric-acid}$ *(not a protein!)*
4. No efficient way of sampling new triples
   useful for data augmentation or negative sampling



*Interpreting the score functions of KGE models as constrained computational graph: **circuits**.*

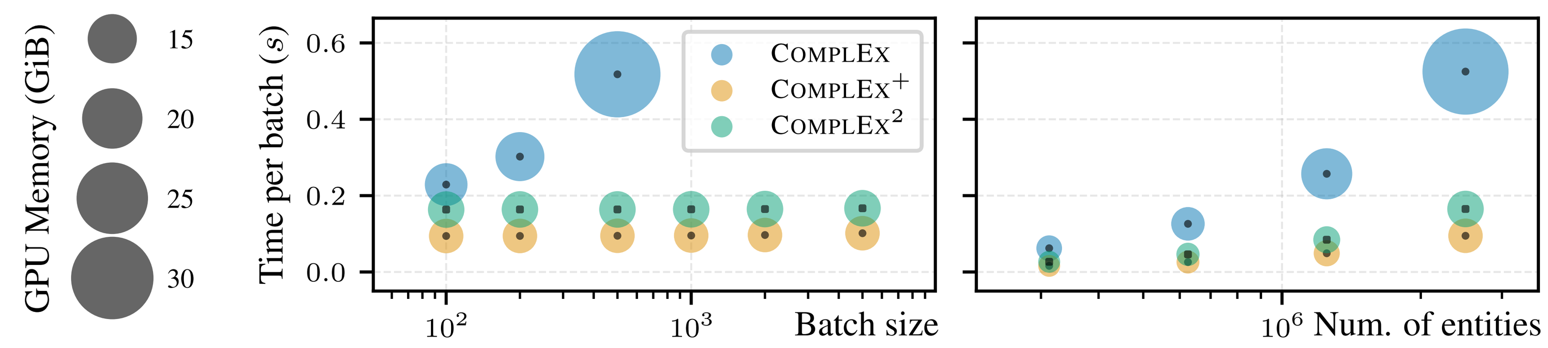
... to Probabilistic Circuits

We cast score functions to *probabilistic circuits* (PCs) [1] ϕ_{pc} without additional memory requirements

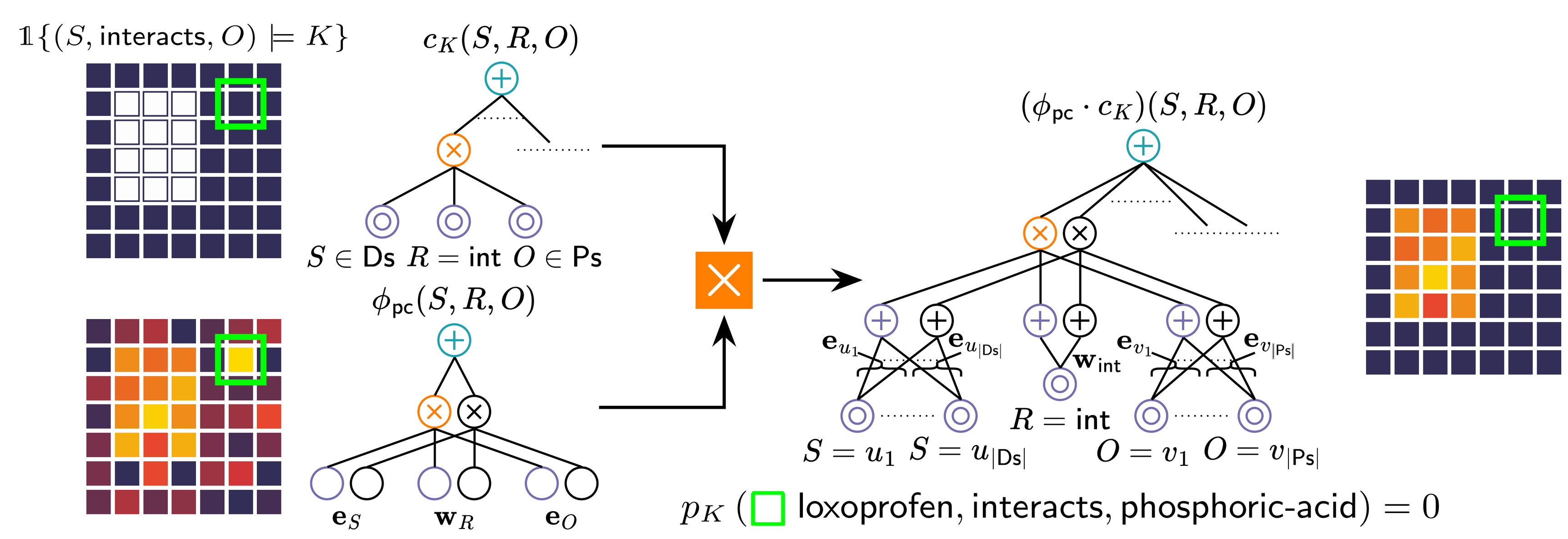
$$p(s,r,o) = \frac{1}{Z}\phi_{\text{pc}}(s,r,o) \quad s.t. \quad \phi_{\text{pc}}(s,r,o)\geq 0$$

- **Non-negative restriction**: make the embeddings and unit activations non-negative [2]
- **Squaring the score function**: square the circuit, without restricting the parameters domain [6], e.g.,

$$\phi_{\text{CP}^2}(s,r,o) = \langle \mathbf{e}_s, \mathbf{w}_r, \mathbf{e}_o\rangle^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} e_{si}e_{sj}w_{ri}w_{rj}e_{oi}e_{oj}$$



Scaling training on KGs with millions of entities.



Integration of domain constraints with guarantees.

The perks

Computing Z and any marginal/conditional probability can be done in time $\mathcal{O}((|\mathcal{E}|+|\mathcal{R}|)\cdot\text{cost}(\phi_{\text{pc}}))$, whilst being more memory efficient at exploiting GPUs

e.g., $Z = \sum_{(s,r,o)}\phi_{\text{CP}}(s,r,o) = \langle \sum_s \mathbf{e}_s, \sum_r \mathbf{w}_r, \sum_o \mathbf{e}_o\rangle$

1. Efficient computation of triple probabilities
   comparable scores across different models and queries
2. More efficient learning by maximum-likelihood and by using classification objectives
3. Propositional logical constraints satisfied *by design*
   e.g., $K := S\in\text{Drugs}\wedge R=\text{interacts}\wedge O\in\text{Proteins}$
   $p_K(\text{loxoprofen}, \text{interacts}, \text{phosphoric-acid}) = 0$
4. Efficient sampling of *high quality* new triples
   via ancestral or inverse transform sampling

Our *Generative KGE Circuits* (GeKCs) are competitive with traditional KGE models, *while coming with the above perks!*

Model	FB15k-237				WN18RR				ogbl-biokg			
	PLL		MLE		PLL		MLE		PLL		MLE	
CP	0.310	(8)	—		**0.105**	(11)	—		0.831	(136)	—	
CP$^+$	0.237	(1)	0.230	(1)	0.027	(1)	0.026	(1)	0.496	(172)	0.501	(142)
CP^2	**0.315**	(8)	0.282	(7)	**0.104**	(23)	0.091	(23)	**0.848**	(66)	0.829	(61)
ComplEx	**0.342**	(36)	—		**0.471**	(16)	—		0.829	(180)	—	
ComplEx$^+$	0.214	(10)	0.205	(5)	0.030	(6)	0.029	(3)	0.503	(245)	0.516	(212)
ComplEx^2	0.334	(10)	0.300	(16)	0.420	(37)	0.391	(19)	**0.858**	(71)	0.840	(59)

References

[1] YooJung Choi, Antonio Vergari, and Guy Van den Broeck. *Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling*. Tech. rep. University of California, Los Angeles (UCLA), 2020.

[2] Alexis de Colnet and Stefan Mengel. "A Compilation of Succinctness Results for Arithmetic Circuits". In: *AAAI* (2021).

[3] Nicolas Hubert et al. "New Strategies for Learning Knowledge Graph Embeddings: The Recommendation Case". In: *EKAW*. Vol. 13514. Lecture Notes in Computer Science. Springer, 2022, pp. 66–80.

[4] Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. "You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings". In: *ICLR*. OpenReview.net, 2020.

[5] Kristina Toutanova and Danqi Chen. "Observed Versus Latent Features for Knowledge Base and Text Inference". In: *3rd Workshop on Continuous Vector Space Models and Their Compositionality*. ACL, 2015.

[6] Antonio Vergari et al. "A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference". In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 13189–13201.