# Your Knowledge Graph Embeddings are Secretly Circuits ... ...and You Should Treat Them as Such


*Lorenzo Loconte*
University of Bari, Italy

Robert Peharz
TU Graz, Austria

Antonio Vergari
University of Edinburgh, UK


## From KGE models ...

A Knowledge Graph Embedding (KGE) model defines a scoring function $\phi$ on triples such that

$$\phi(s,p,o) \propto \log \Pr(s,p,o)$$

thus being an Energy-Based Model (EBM). Popular KGE models such as CP, RESCAL and TuckER define $\phi$ as

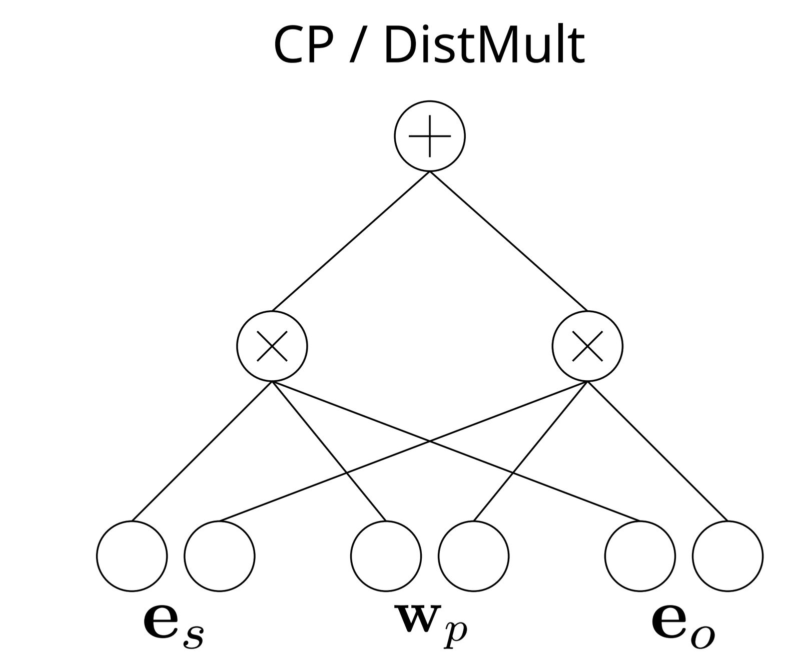
$$\phi_{\text{CP}}(s,p,o) = \langle \mathbf{e}_s, \mathbf{w}_p, \mathbf{e}_o \rangle = \sum\nolimits_{i=1}^{R} e_{si} w_{pi} e_{oi}$$

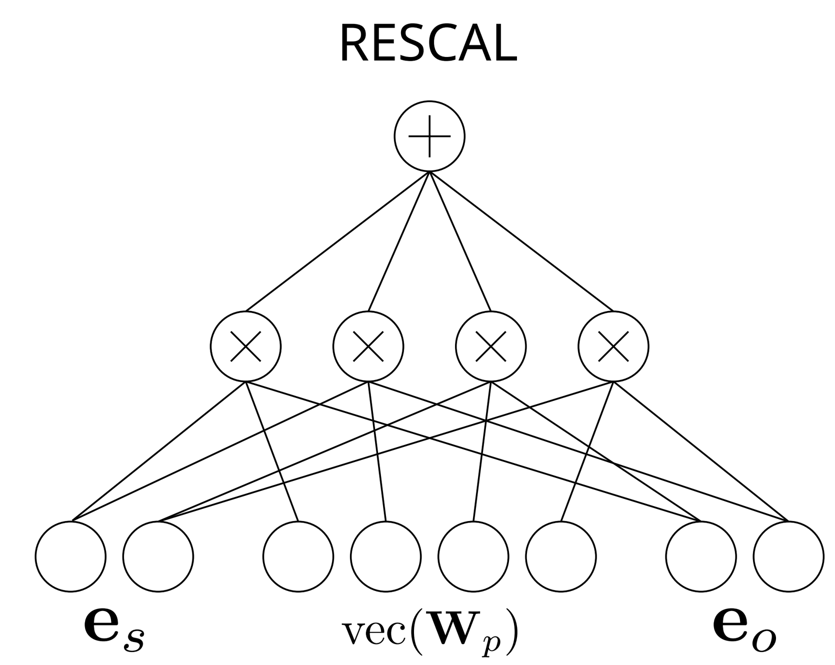
$$\phi_{\text{RESCAL}}(s,p,o) = \mathbf{e}_s^T \mathbf{W}_p \mathbf{e}_o$$

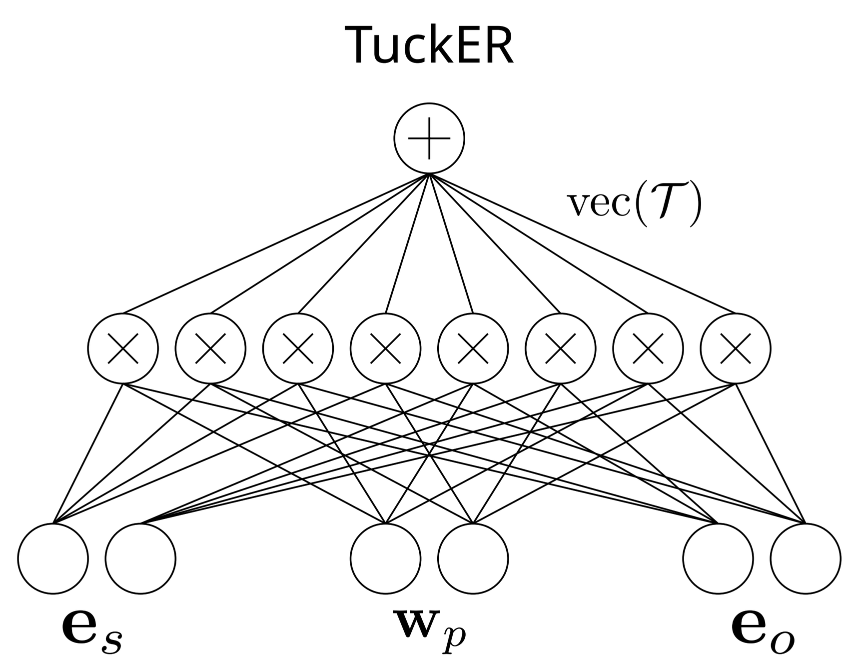
$$\phi_{\text{TuckER}}(s,p,o) = \mathcal{T} \times_1 \mathbf{e}_s \times_2 \mathbf{w}_p \times_3 \mathbf{e}_o$$

KGE models have several shortcomings:

1. Exact probabilistic inference is impractical
*e.g.* $\mathcal{Z} = \sum_{s\in\mathcal{E}} \sum_{p\in\mathcal{R}} \sum_{o\in\mathcal{E}} \exp \phi(s,p,o)$
*requires* $10^{19}$ *evaluations of* $\phi$ *for Freebase [3]*
2. No efficient way of sampling new triples
3. Learning by maximum-likelihood is not supported
4. No principled probabilistic complex query answering
*e.g. "Which drugs interact with proteins associated with the diseases* $d_1$ *or* $d_2$ *?"*



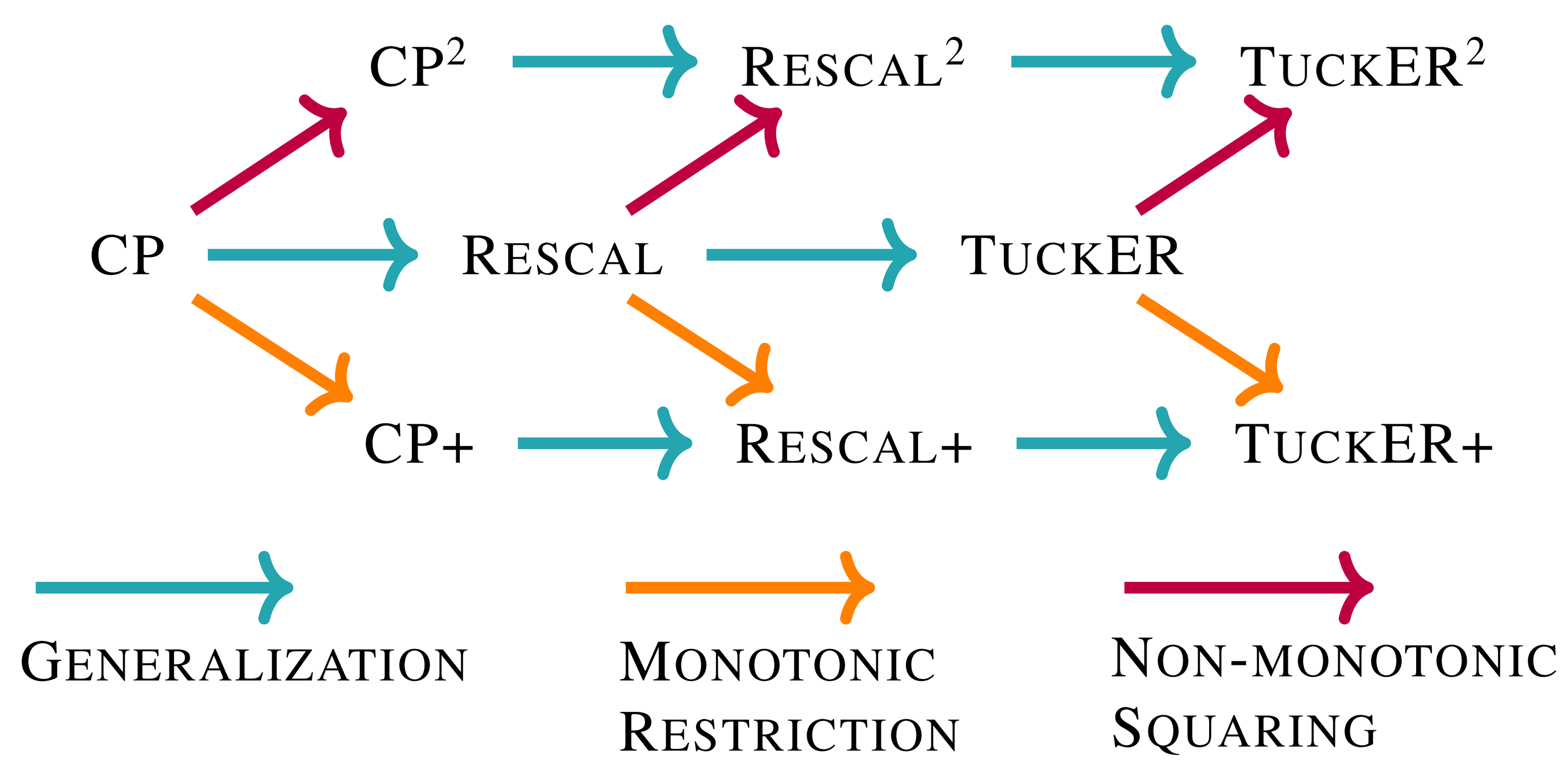

CP, DistMult, RESCAL, TuckER encode tensor factorizations that define scoring functions that can be readily represented as tractable computational graphs with input, product and weighted sum units, called *circuits*.

## ...to probabilistic circuits

KGE models based on tensor factorizations can be cast to *Probabilistic Circuits* [1] modeling a probability distribution over triples in two ways:

**Monotonic Restriction** $\rightarrow$ restrict the embeddings and additional parameters to be non-negative [2];

**Non-monotonic Squaring** $\rightarrow$ square non-monotonic circuits, without additional constraints [4].




## The perks

By doing so, we obtain tractable generative KGE models encoding a probability distribution over triples. This enables:

1. Efficient and exact probabilistic inference
*e.g. computing* $\mathcal{Z}$ *in a single feedforward pass*
2. Efficient way of sampling of new triples
*via ancestral or inverse transform sampling*
3. Learning by Maximum Likelihood Estimation and with composite objectives *e.g.* $\mathcal{L}_{\text{1vsALL}} + \mathcal{L}_{\text{MLE}}$
4. Principled and probabilistic way to answer complex queries

Preliminary experiments with CP show that its generative counterparts can be as expressive (but come with the above perks!)

| Dataset | Model | 1vsAll | | MLE | | 1vsAll+MLE | |
|---|---|---|---|---|---|---|---|
| | | **MRR** | **Hits@1** | **MRR** | **Hits@1** | **MRR** | **Hits@1** |
| Nations | CP | 0.793 ±0.004 | 0.679 ±0.008 | — | — | — | — |
| | CP+ | 0.801 ±0.004 | 0.695 ±0.007 | 0.788 ±0.004 | 0.683 ±0.007 | 0.795 ±0.007 | 0.692 ±0.012 |
| | $CP^2$ | 0.796 ±0.003 | 0.700 ±0.004 | 0.801 ±0.004 | 0.698 ±0.007 | **0.809** ±0.001 | **0.706** ±0.001 |
| UMLS | CP | **0.954** ±0.003 | **0.916** ±0.006 | — | — | — | — |
| | CP+ | 0.856 ±0.005 | 0.762 ±0.008 | 0.850 ±0.004 | 0.750 ±0.007 | 0.856 ±0.005 | 0.761 ±0.009 |
| | $CP^2$ | 0.927 ±0.001 | 0.881 ±0.002 | 0.898 ±0.002 | 0.810 ±0.004 | 0.897 ±0.001 | 0.808 ±0.002 |
| Kinship | CP | 0.858 ±0.002 | 0.773 ±0.003 | — | — | — | — |
| | CP+ | 0.725 ±0.002 | 0.603 ±0.004 | 0.735 ±0.003 | 0.615 ±0.004 | 0.728 ±0.005 | 0.605 ±0.007 |
| | $CP^2$ | 0.872 ±0.002 | 0.800 ±0.004 | **0.891** ±0.002 | **0.829** ±0.003 | **0.889** ±0.001 | **0.827** ±0.002 |

### References


[1] YooJung Choi, Antonio Vergari, and Guy Van den Broeck. *Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling*. Tech. rep. 2020.

[2] Alexis de Colnet and Stefan Mengel. "A Compilation of Succinctness Results for Arithmetic Circuits". In: *arXiv preprint arXiv:2110.13014* (2021).

[3] Maximilian Nickel et al. "A Review of Relational Machine Learning for Knowledge Graphs". In: *IEEE* 104.1 (2016), pp. 11–33.

[4] Antonio Vergari et al. "A Compositional Atlas of Tractable Circuit Operations: From Simple Transformations to Complex Information-Theoretic Queries". In: *arXiv preprint arXiv:2102.06137* (2021).