TN-SHAP-G

Farzaneh Heidari, Guillaume Rabusseau

Université de Montréal · Mila · CIFAR

Shapley Values & Explainability

The Shapley value assigns each feature a fair share of the prediction from its **average contribution over all subsets.**

$$\phi(v) = \sum_{S \subseteq V \setminus \{v\}} \frac{|S|!\,(n-|S|-1)!}{n!} \left[\nu(S \cup \{v\}) - \nu(S)\right].$$

Computing Shapley is exponential.
TN-SHAP-G makes it tractable!

Coalition Tensor & Multilinear Extension

$$\mathcal{T} \in \mathbb{R}^{2 \times \cdots \times 2}$$

Coalition values arranged as an (n)-way tensor.

Theorem: Multilinear extension of games as tensor contraction

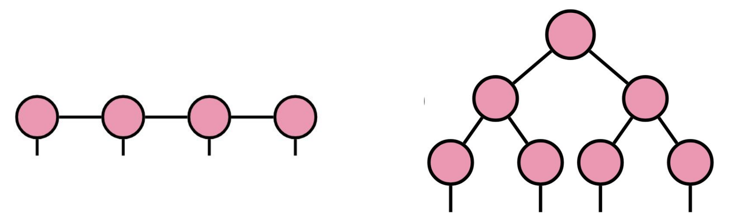
$$\mathcal{T}_{s_1,\dots,s_n} := \nu\big(S(s)\big) \quad s \in \{0,1\}^n$$

Continuous extension of the binary mask $z \in [0,1]^n$

$$\widetilde{\nu}(z) = \sum_{s \in \{0,1\}^n} \mathcal{T}_s \prod_{v=1}^{n} \mathbf{b}_v \qquad \widetilde{\nu} : [0,1]^n \to \mathbb{R} \qquad \mathbf{b}_v(z_v) := \begin{bmatrix} 1 - z_v \\ z_v \end{bmatrix}$$

Multilinear extension of the game

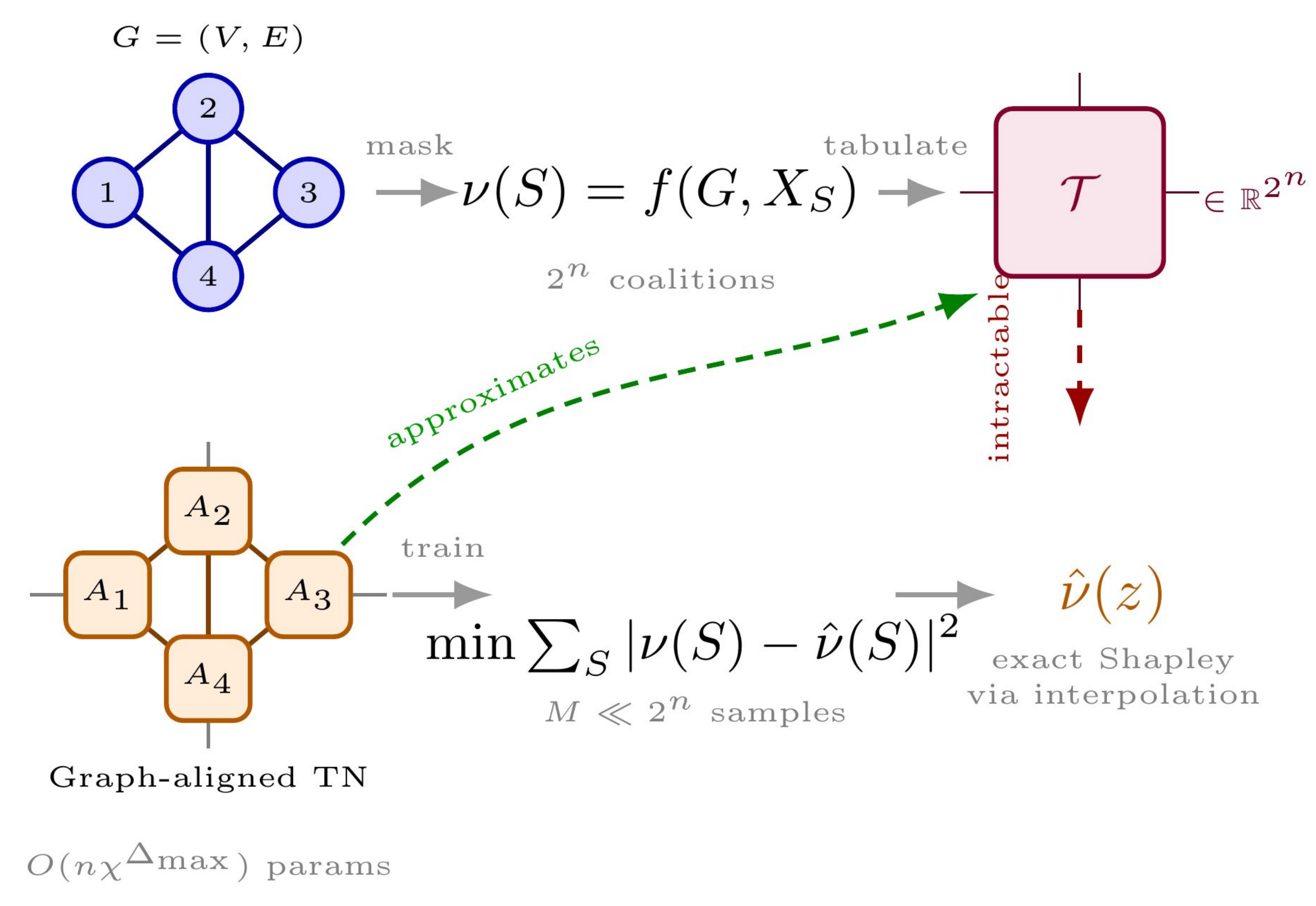
then: $\phi(u) = \int_0^1 \frac{\partial \widetilde{\nu}}{\partial z_u}(t\mathbf{1})\,dt$



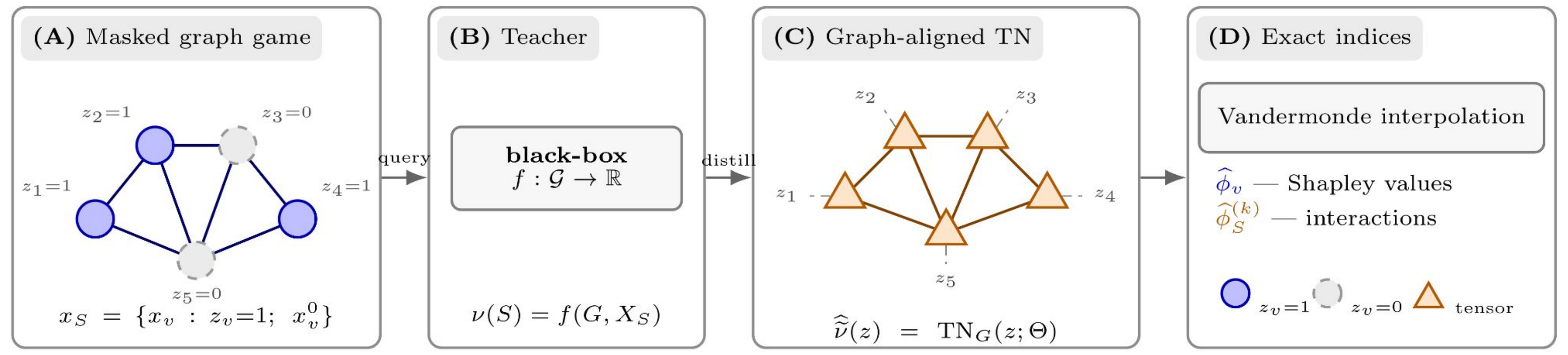
TT · Tensor Tree

Low-rank tensor networks compress the coalition tensor using small factors.

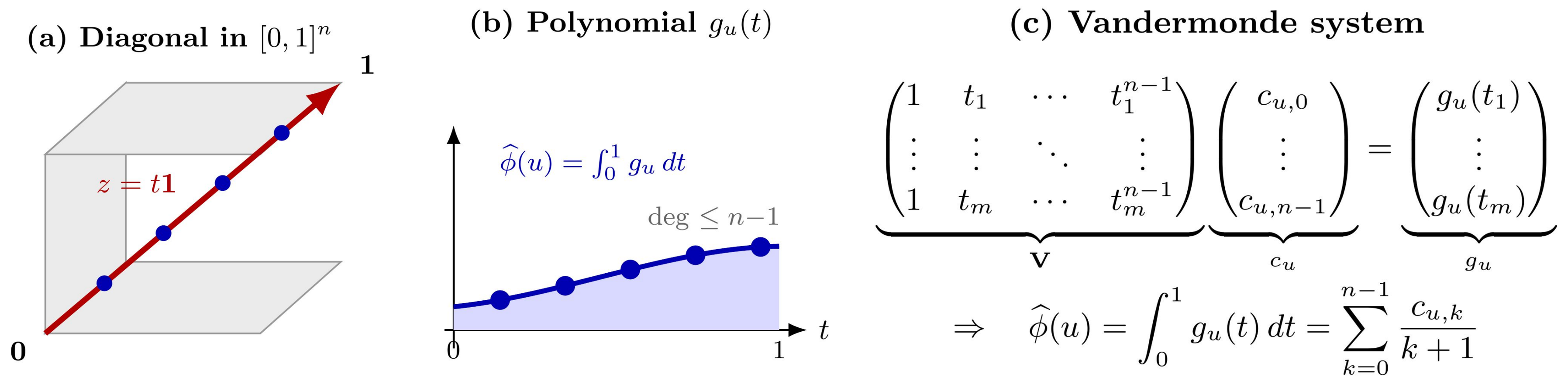
Graph-structured Tensor Network Surrogates for Shapley Values and Interactions



We query the black-box model to obtain a graph-aligned TN surrogate. Shapley values and interactions are then computed deterministically from the surrogate.

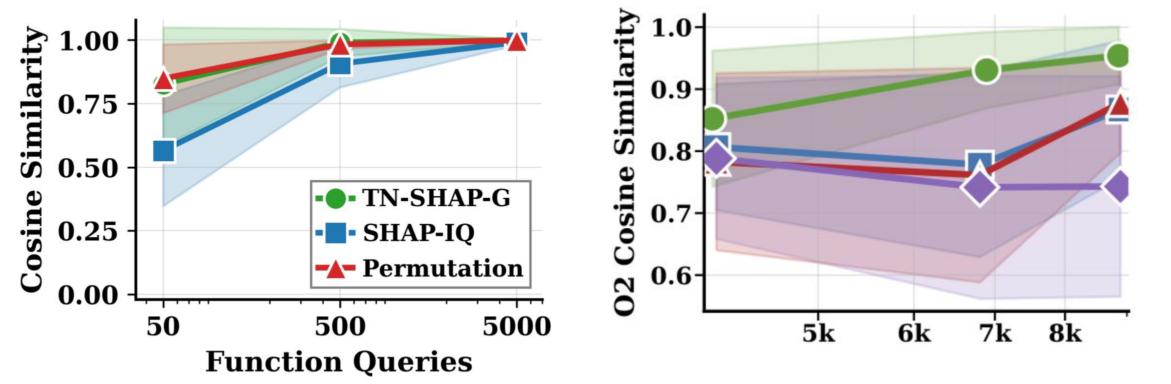


Vandermonde Interpolation

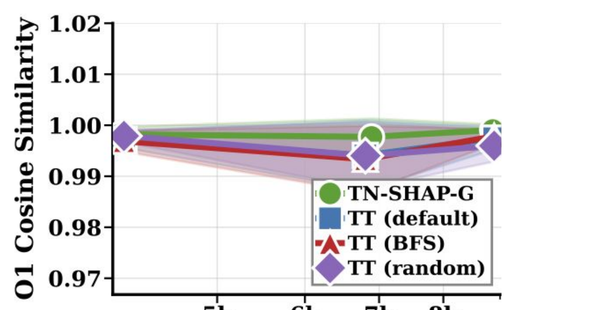


Deterministic Shapley Value Computation on a TN Surrogate via Diagonal Interpolation

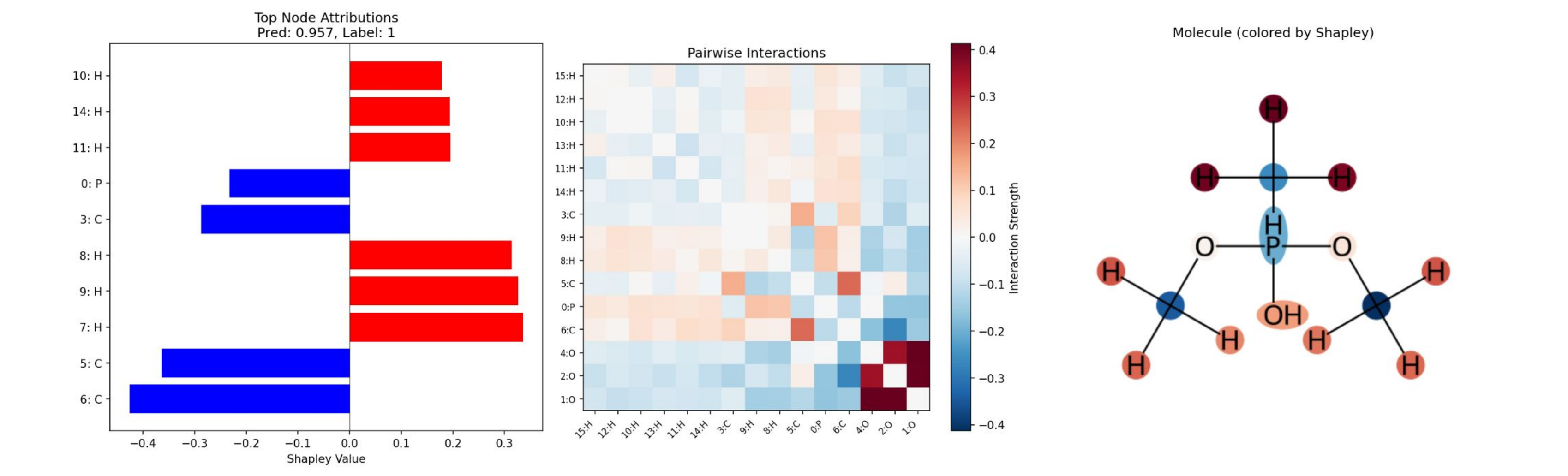
Experimental Results



Comparison of TN-SHAP-G with baselines



TN topology ablation



Qualitative example of TN-SHAP-G on MUTAG. The most important atoms are faithfully identified.