TN-SHAP
Tractable Shapley Values & Interactions via Tensor Networks

Farzaneh Heidari
Chao Li
Guillaume Rabusseau

Mila AIP Université de Montréal CIFAR

Shapley Values & Explainability

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

$$\Phi_i(x) = \sum_{s=0}^{n-1} \frac{s!(n-s-1)!}{n!} \sum_{\substack{C \subseteq N\setminus\{i\} \\ |C|=s}} \left[v(x, C \cup \{i\}) - v(x, C)\right].$$

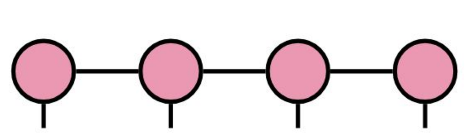
An exponential sum!

Coalition Tensor

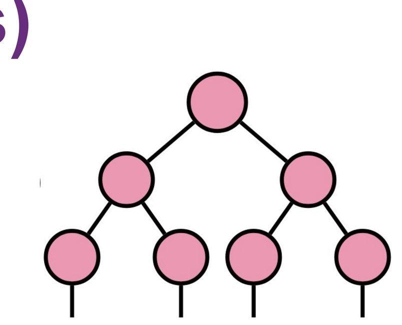
Coalition values can be arranged as an (n−1)-way tensor.

$$\mathcal{T} \in \mathbb{R}^{\overbrace{2\times\cdots\times 2}^{\text{(n-1) times}}}$$

Tensor Networks (TNs)

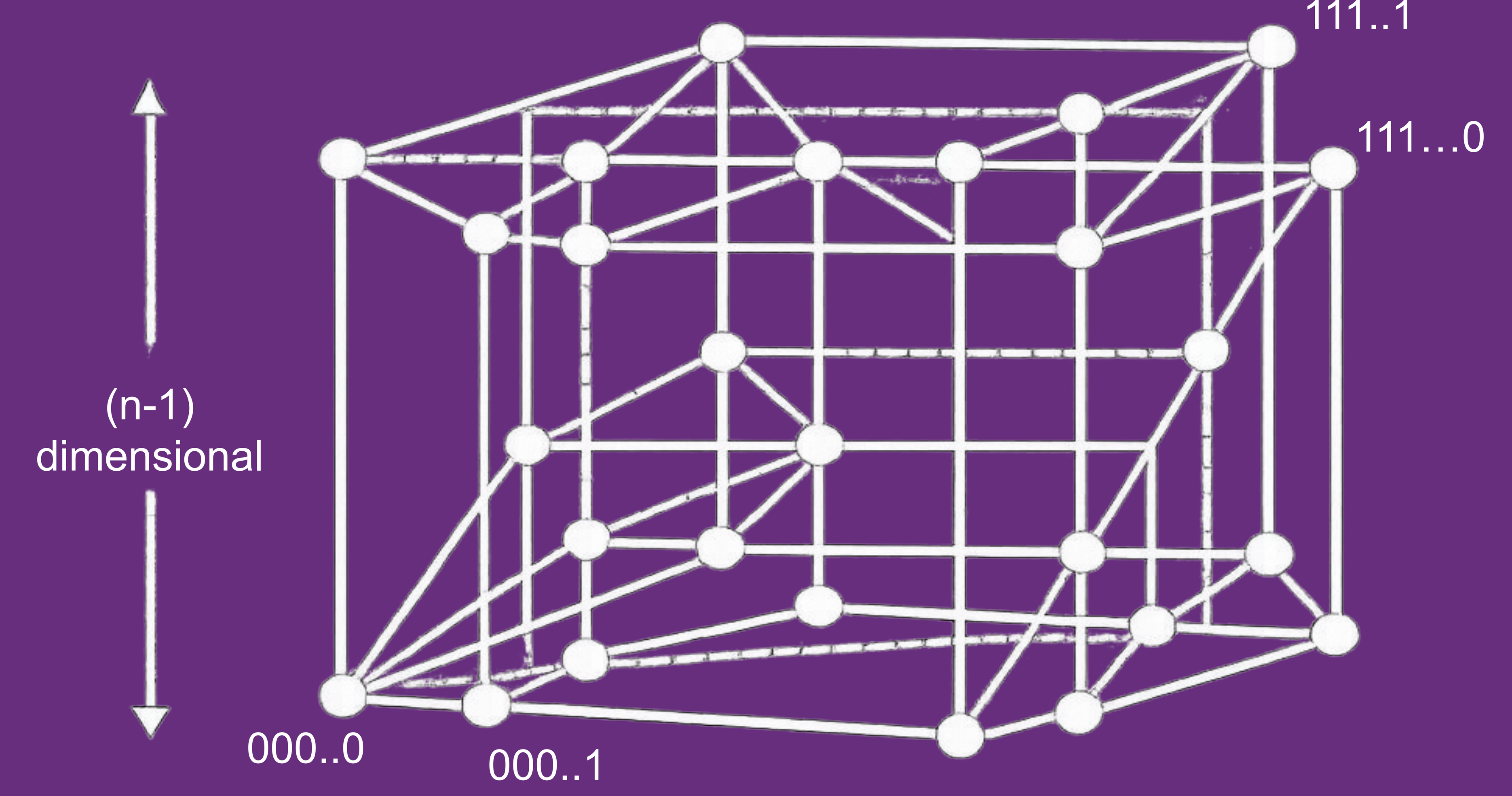


TT

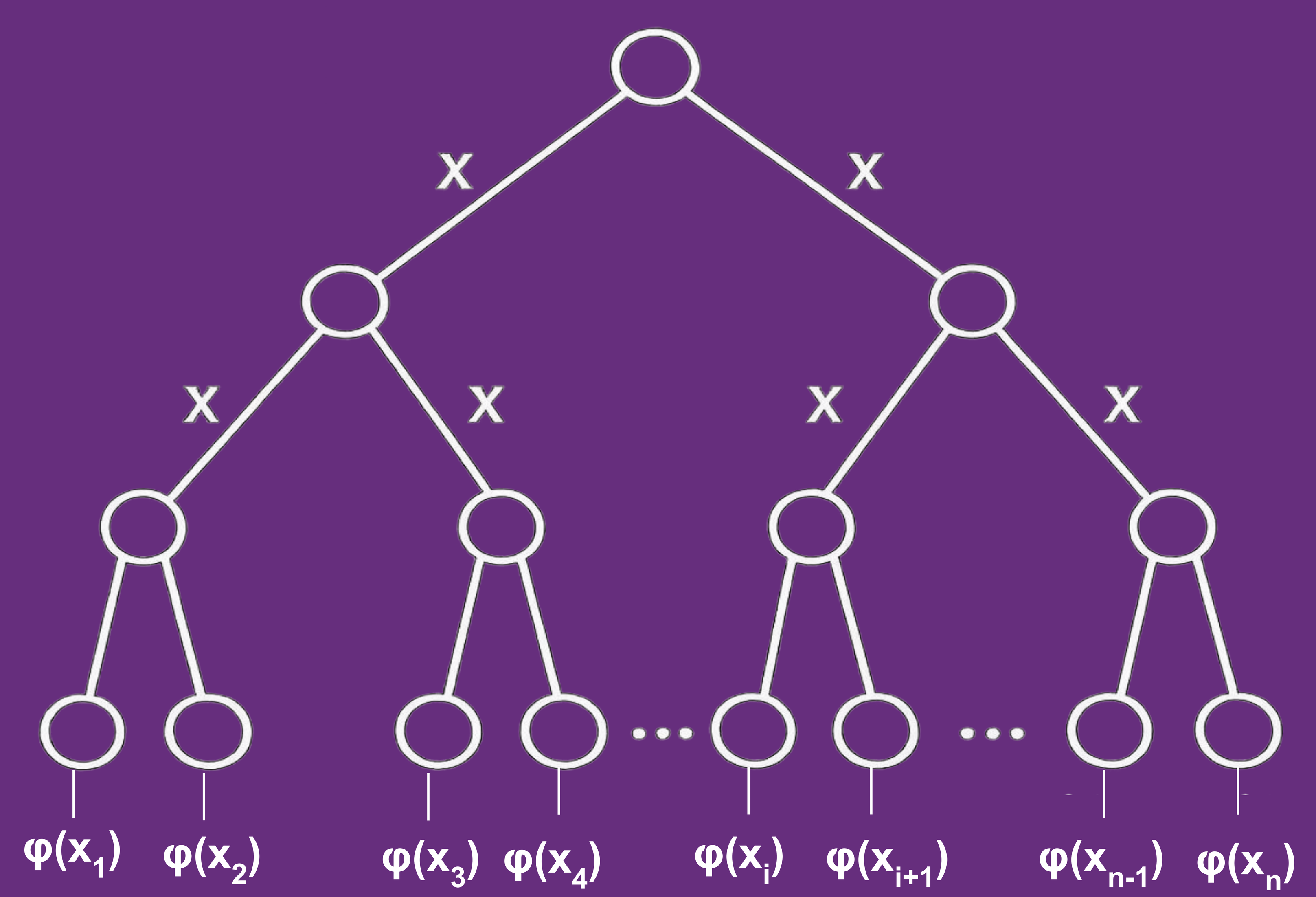


Tensor Tree

Low-rank tensor networks compress exponential tensors using small factors.



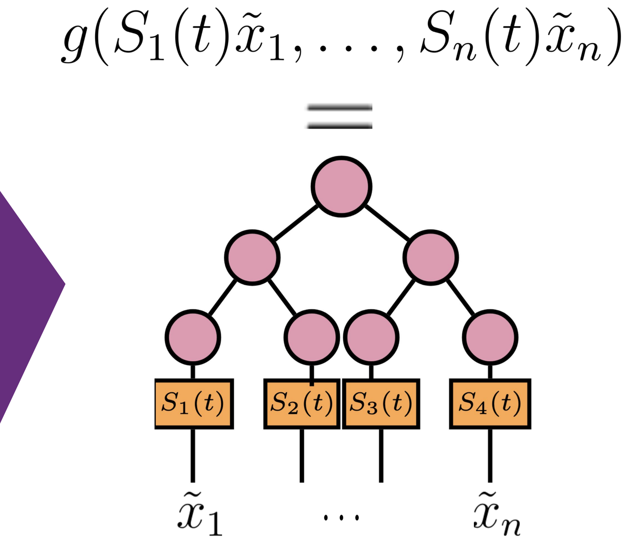
TN-SHAP
turns exponential coalitions into structured TN probes



Method

$$\tilde{x}_i = [x_i, 1]^\top$$

$$S_r(t) = \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}$$



$$G_i(t; x) := g(\mathrm{on}_i, \{S_r(t)\}_{r\neq i}) - g(\mathrm{off}_i, \{S_r(t)\}_{r\neq i})$$

aggregates marginal contributions across coalition sizes

$$G_i(t; x) = \sum_{s=0}^{n-1} m_s^{(i)}(x) \cdot t^s$$

Vandermonde solve

$$\begin{bmatrix} 1 & t_0 & \cdots & t_0^{n-1} \\ 1 & t_1 & \cdots & t_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_{n-1} & \cdots & t_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} m_0^{(i)} \\ m_1^{(i)} \\ \vdots \\ m_{n-1}^{(i)} \end{bmatrix} = \begin{bmatrix} G_i(t_0; x) \\ G_i(t_1; x) \\ \vdots \\ G_i(t_{n-1}; x) \end{bmatrix}$$

$$\Phi_i(x) = \sum_{s=0}^{n-1} \frac{s!(n-s-1)!}{n!} m_s^{(i)}(x)$$

Shapley computation in:

$$\mathbf{O(n \cdot poly(\chi) + n^2)}$$

Experiments

25–1000× faster, same fidelity.