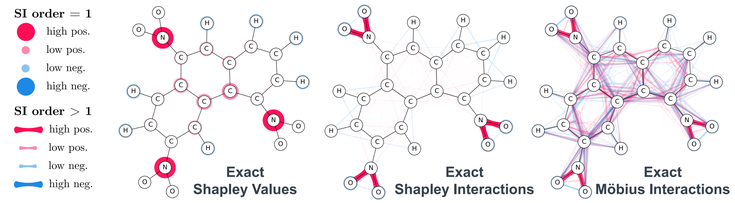
Exact Computation of Any-Order Shapley Interactions for Graph Neural Networks

Maximilian Muschalik*, Fabian Fumagalli*, Paolo Frazzetto, Janine Strotherm, Luca Hermes, Alessandro Sperduti, Eyke Hüllermeier, and Barbara Hammer

Paper: Open Review

Contribution Summary



TL;DR: Most Möbius interactions are zero for *graph-level prediction* tasks with GNNs, allowing for exact computation of Shapley Values and Interactions.

- The *molecule* on the left has *30 atoms* (graph with 30 nodes). Computing **exact** explanations (no approximation) requires $2^{30} \approx 1{,}000{,}000{,}000$ model calls.
- **GraphSHAP-IQ needs only 7,693.** 🚀

Background

A **Graph Neural Network** (GNN) makes prediction for a graph g:

$$f_g(\mathbf{X}) := \sigma\big(\Psi(\{\{f_i(\mathbf{X}) \mid v_i \in V\}\})\big) \quad \text{with} \quad \begin{matrix}\sigma\text{: output layer} \\ \Psi\text{: pooling function}\end{matrix}$$

Möbius Interactions $m : \mathcal{P}(N) \to \mathbb{R}$ are the basis of explanations:

$$m(S) := \sum_{T \subseteq S} (-1)^{|S|-|T|} \nu(T) \quad \text{and they recover} \quad \nu(T) = \sum_{S \subseteq T} m(S)$$

Shapley Values and **Interactions** summarize/aggregate the Möbius interactions into lower-order explanations.

- *Shapley Values* measure feature attribution of nodes
- *Shapley Interactions* measure synergy between nodes

References

For full references, see our accompanying paper.

Shapley, L. S. (1953). A Value for n-Person Games. Contributions to the Theory of Games, Volume II, pages 307–318.
Grabisch and Roubens (1999). An Axiomatic Approach to the Concept of Interaction Among Players in Cooperative Games. Int. J. Game Theory 28(4):547–565
Bordt and von Luxburg (2023). From Shapley Values to Generalized Additive Models and Back. AISTATS'23
Lundberg et al. (2017). A Unified Approach to Interpreting Model Predictions. NeurIPS'17
Fumagalli et al. (2023). SHAP-IQ: Unified Approximation of any-order Shapley Interactions. NeurIPS'23
Ye et al. (2023). SAME: uncovering GNN black box with structure-aware shapley-based multipiece explanations. NeurIPS'23
Amara et al. (2022) Graphframex: Towards systematic evaluation of explainability methods for graph neural networks. LoG'22

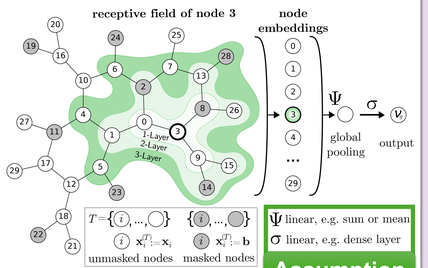
GraphSHAP-IQ

The Graph Game transforms a GNN for graph-level prediction into a *cooperative game* by masking nodes:

Game: $\nu_g(T) := f_{g,\hat{y}}(\mathbf{X}^{(T)})$

with $\mathbf{X}^{(T)} := (\mathbf{x}_1^{(T)}, ..., \mathbf{x}_n^{(T)})^t \in \mathbb{R}^{n \times d_0}$

and $\mathbf{x}_i^{(T)} := \begin{cases} \mathbf{x}_i & \text{if } i \in T, \\ \mathbf{b} & \text{if } i \notin T, \end{cases}$



Ψ linear, e.g. sum or mean
σ linear, e.g. dense layer

Assumption

See the paper for deep redout!

Node-Level Results:

Theorem 3.3

"only masking within the receptive fields matter"

Graph-Level Results:

Proposition 3.6

"most Möbius interactions are zero for graph-level GNNS"

Algorithm 1 GraphSHAP-IQ

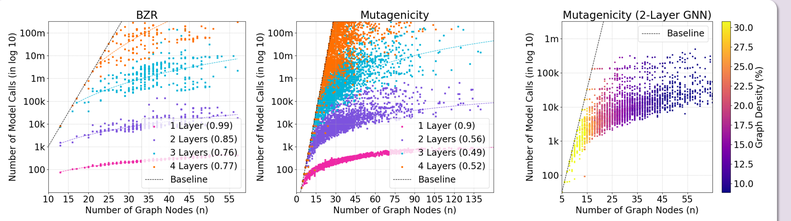
Require: Graph $g = (V, E, \mathbf{X})$, ℓ-Layer GNN f_g, SI order k.

$\mathcal{I} \leftarrow \bigcup_{i \in N} \mathcal{P}(\mathcal{N}_i^{(\ell)})$
$\nu_g \leftarrow [f_g(\mathbf{X}^{(T)})]_{T \in \mathcal{I}}$
$m_g \leftarrow [\mathtt{MI}(\nu_g, S)]_{S \in \mathcal{I}}$
$\Phi_k \leftarrow \mathtt{MI\,to\,SI}(m_g, k)$
return MIs m_g, SIs Φ_k

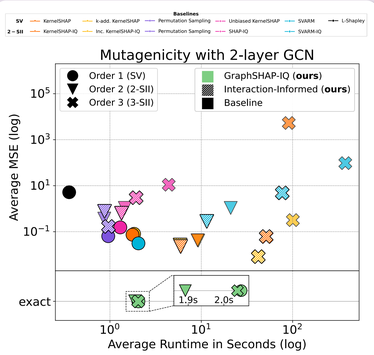
Theorem 3.7

"complexity is bounded by the receptive field"

Empirical Results



- GrapSHAP-IQ scales **near-linearly** with graph size making exact computations feasible
- less *dense* graphs are easier to compute

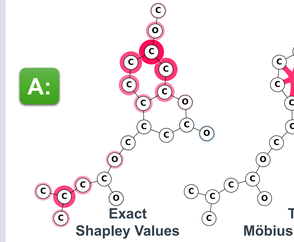


- GraphSHAP-IQ's **runtime** is independent of interaction order
- **interaction-informed** baselines (knowledge about which Möbius interactions are zero) **improve** approximation performance

Example Detecting Benzene Rings:

Q: *Why is this larger molecule a Benzene?*

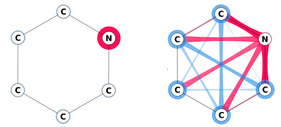
A:



Exact Shapley Values — Top-1 Möbius Interaction

Q: *Why is Pyridyne **no** Benzene?*

A:



1st order Shapley Values — 2nd order Shapley Interactions

ICLR

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

UNIVERSITÄT BIELEFELD

UNIVERSITÀ DEGLI STUDI DI PADOVA

TRR 318

MCML Munich Center for Machine Learning

Soon™ in shapiq!

Check out our other work: