Probing Graph Representations of Molecules

Mohammad Sadegh Akhondzadeh — mohammad.akhondzadeh@cispa.de
Vijay Lingam — vijay.lingam@cispa.de
Aleksandar Bojchevski — bojchevski@cispa.de

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Overview

- We study what information is captured in the learned representations of molecules via probing
- Graph transformers tend to learn richer representations
- Randomly initialized models are surprisingly good
- Probing provides model level explanation

Theoretical v.s. Practical Expressivity

Proofs determining the expressiveness power of GNNs do not consider node features. (Anonymous setting)

A **theoretically more expressive** GNN does not guarantee that it will learn **more expressive and better representation**

Research Question

Can we discern the information encoded in the learned representation of graph-based neural network?

Setup

We use pre-trained models on HOMO-LUMO gap using PCQM4Mv2 dataset.

Freeze the model parameters and generate representations

Apply probing framework on these representations

Probing Tasks

Atom counting: #Carbon, #Oxygen, #Nitrogen

Meaningful substructures (Functional groups): Arom. rings, Benzene, etc.

3D Properties: Asphericity, Radius of Gyration, etc.

High level Properties (transferability): Toxicity, HIV, etc.

Strategy 1: Probing with Linear Classifiers

Can you **predict** the property from the frozen representations?

Probing dataset $D = \{(r_i, p_i)\}_{i=1}^N$

Representation — Property

Probing performance ≈ Extractability (Usability)

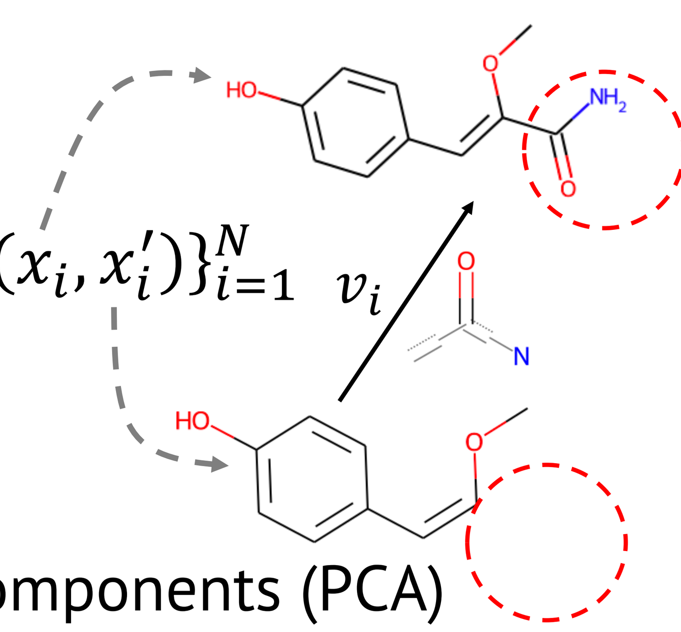
Strategy 2: Bayesian Probing

(More details in paper)

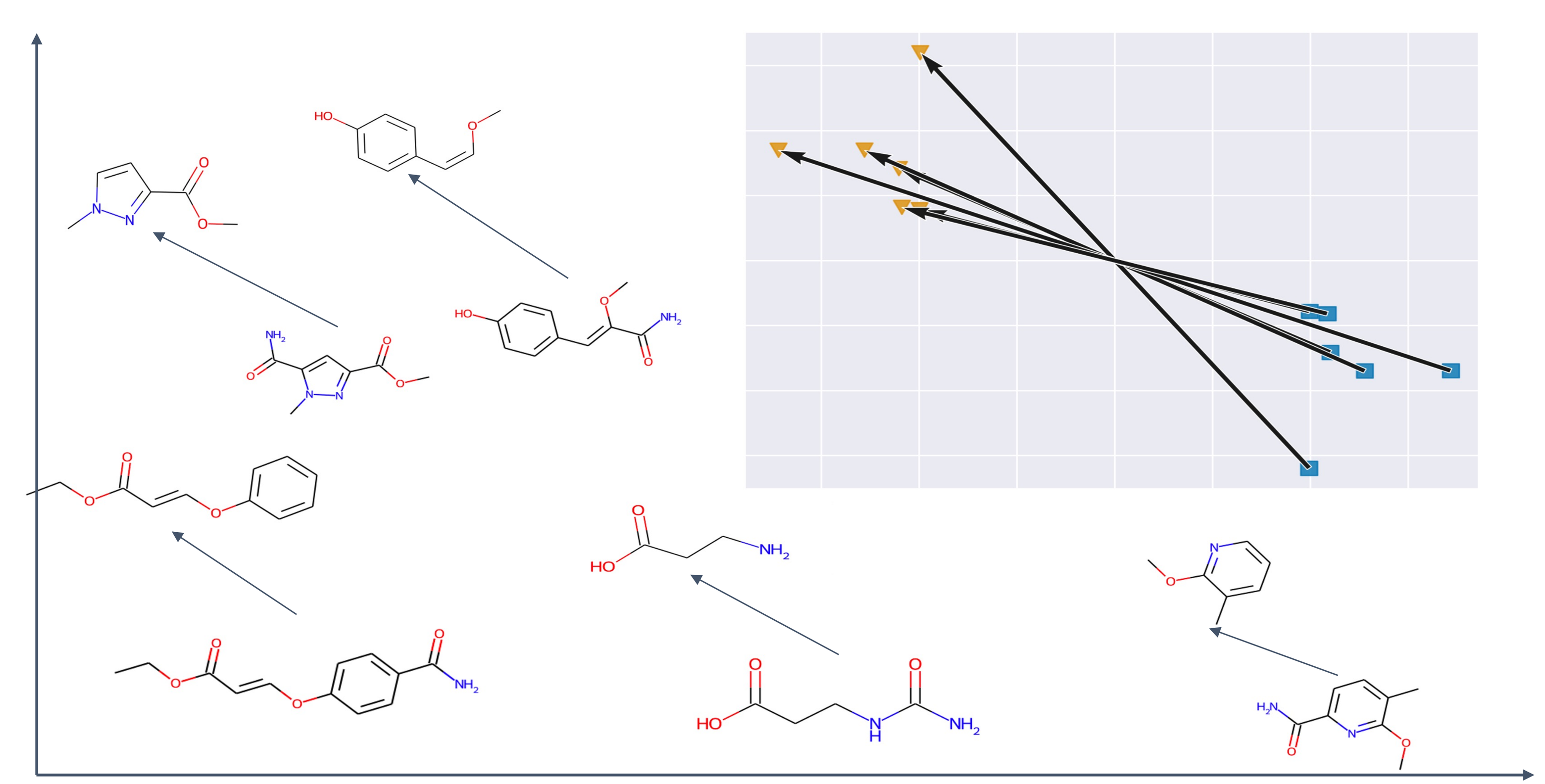
Strategy 3: Pairwise Probing

We construct pairs of molecules that differ only in the property of interest

$D = \{(x_i, x_i')\}_{i=1}^N \quad v_i$



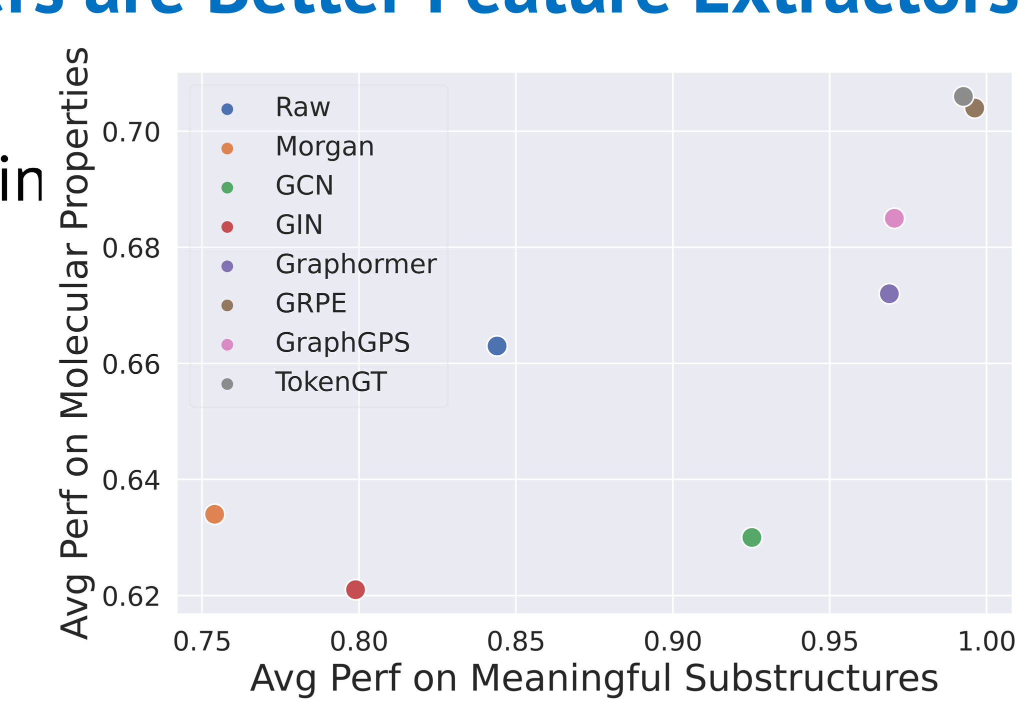
Project onto the first two principal components (PCA)



Graph Transformers are Better Feature Extractors

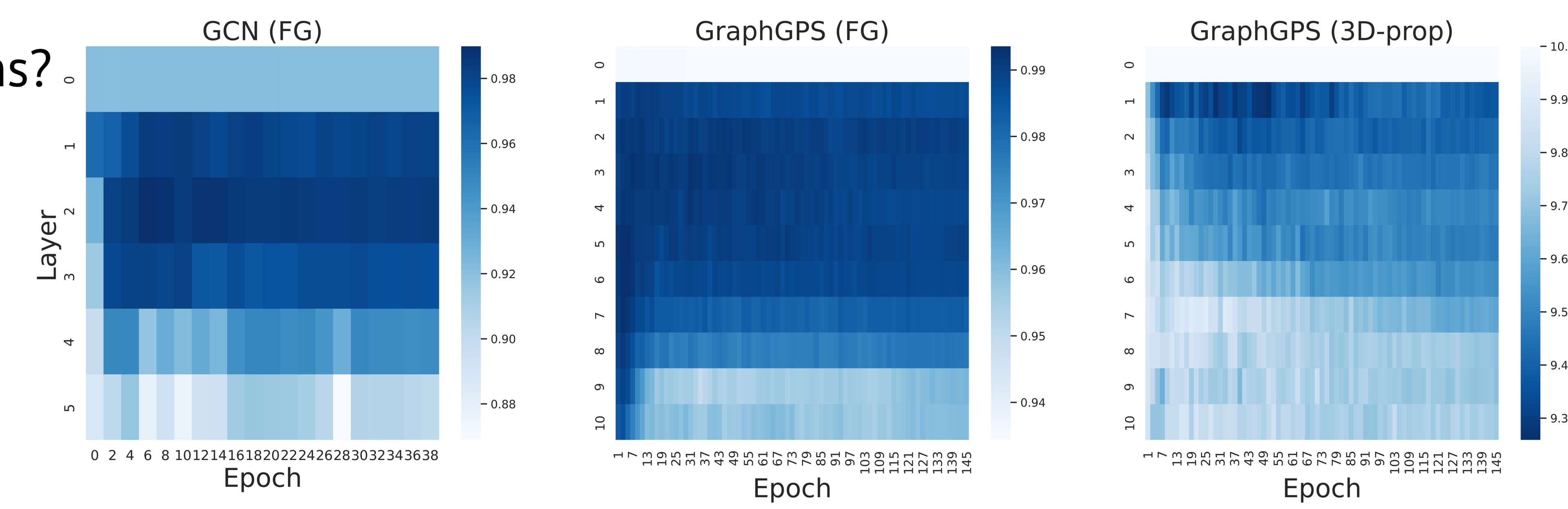
Input features contain a lot of information

Transformer-based models perform better on average



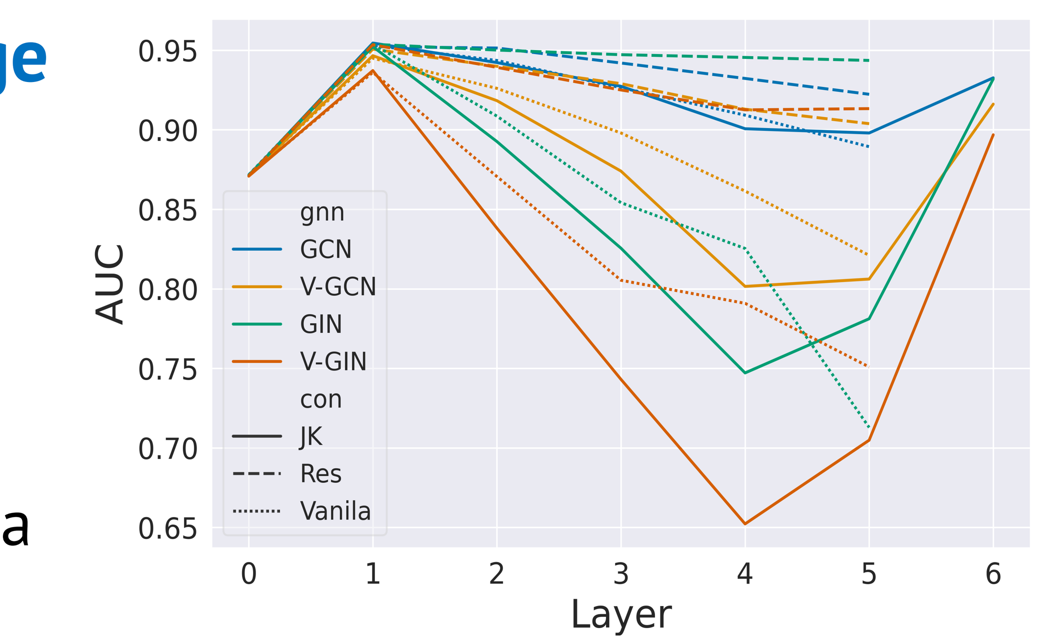
Strong correlation is observed between the performance of probing on detecting substructures & MoleculeNet tasks

Randomly Init. Models Capture Functional Groups



Residual Connections and Jumping Knowledge Preserve Linear Separability

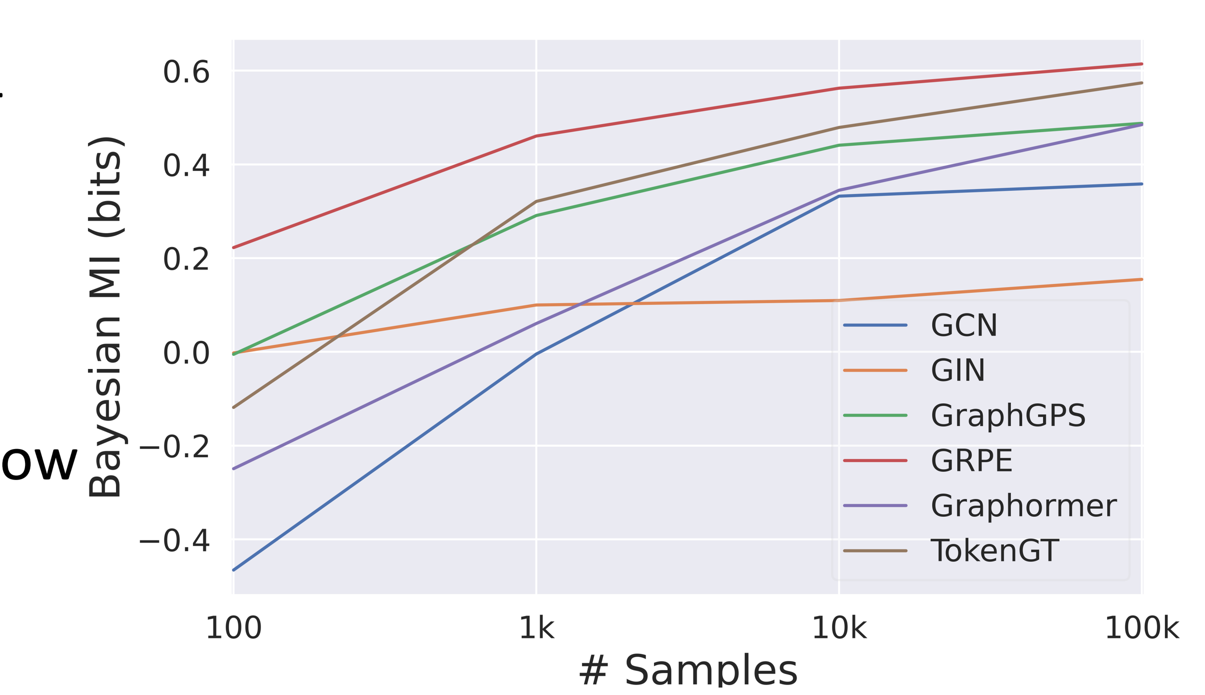
One layer of message-passing is enough for detecting the existence of a sub-structure



Measuring the Ease of Extracting Information by BMI

For the extremely low-data scenario GIN performs surprisingly better

Graph transformers show higher gains with increase in the size of probing dataset



Alignment Along Identified Sub-structures Direction