What we have learned from GNN?

Graph Embedding: from math to model

What should we do next?

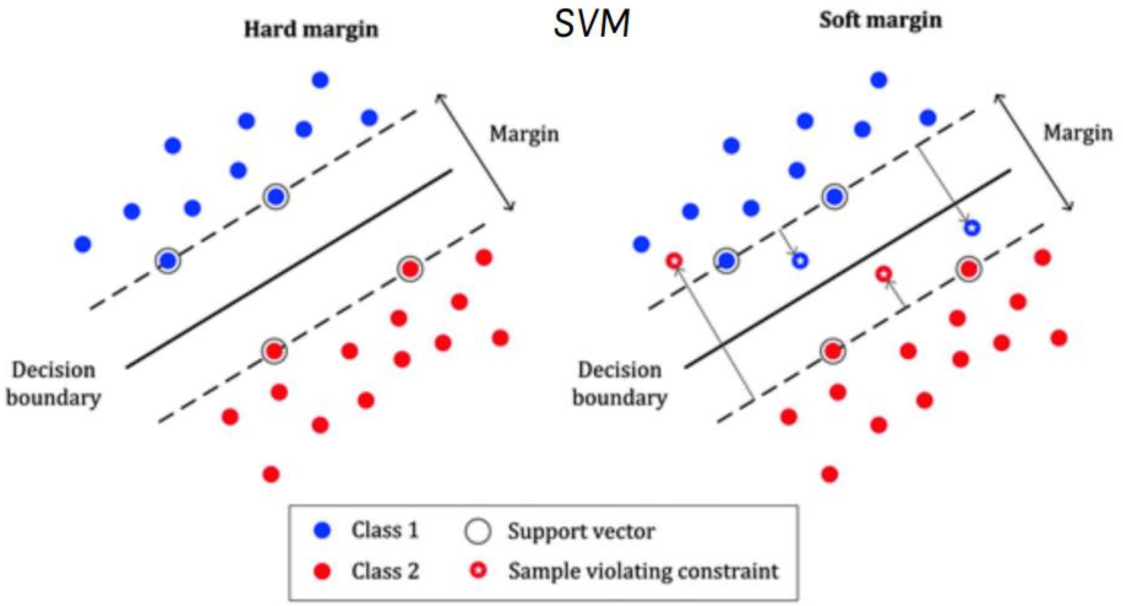
Beyond the model, focus on the paper.

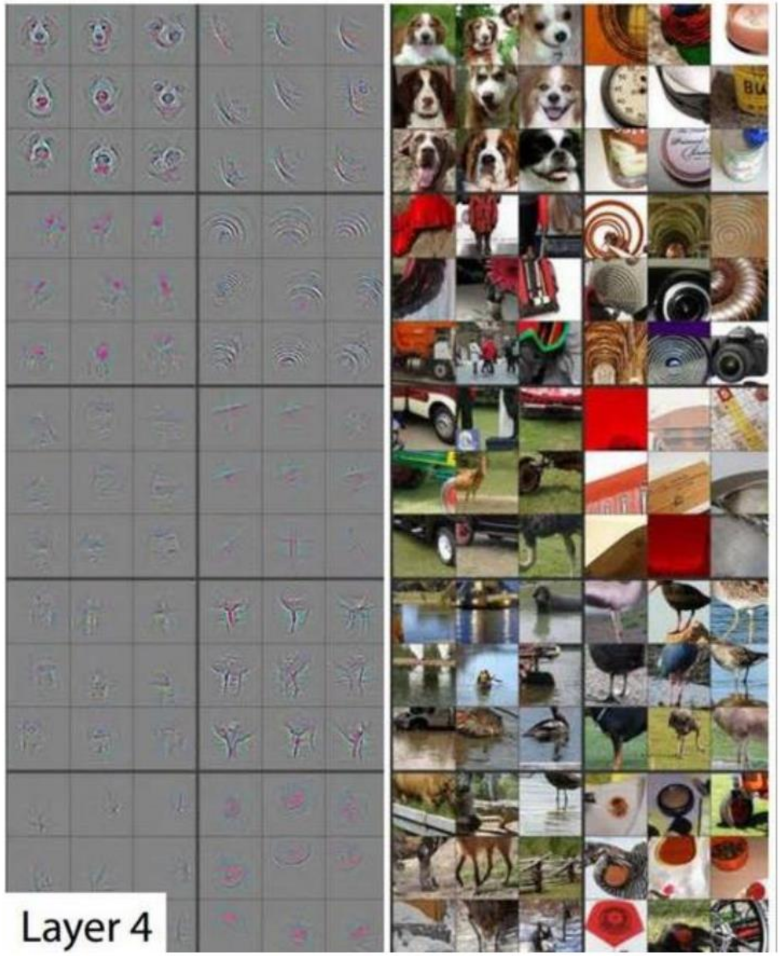
What is the difference between human and NN?

Human: from the whole picture to detail

Deep Network: from detail to the whole picture

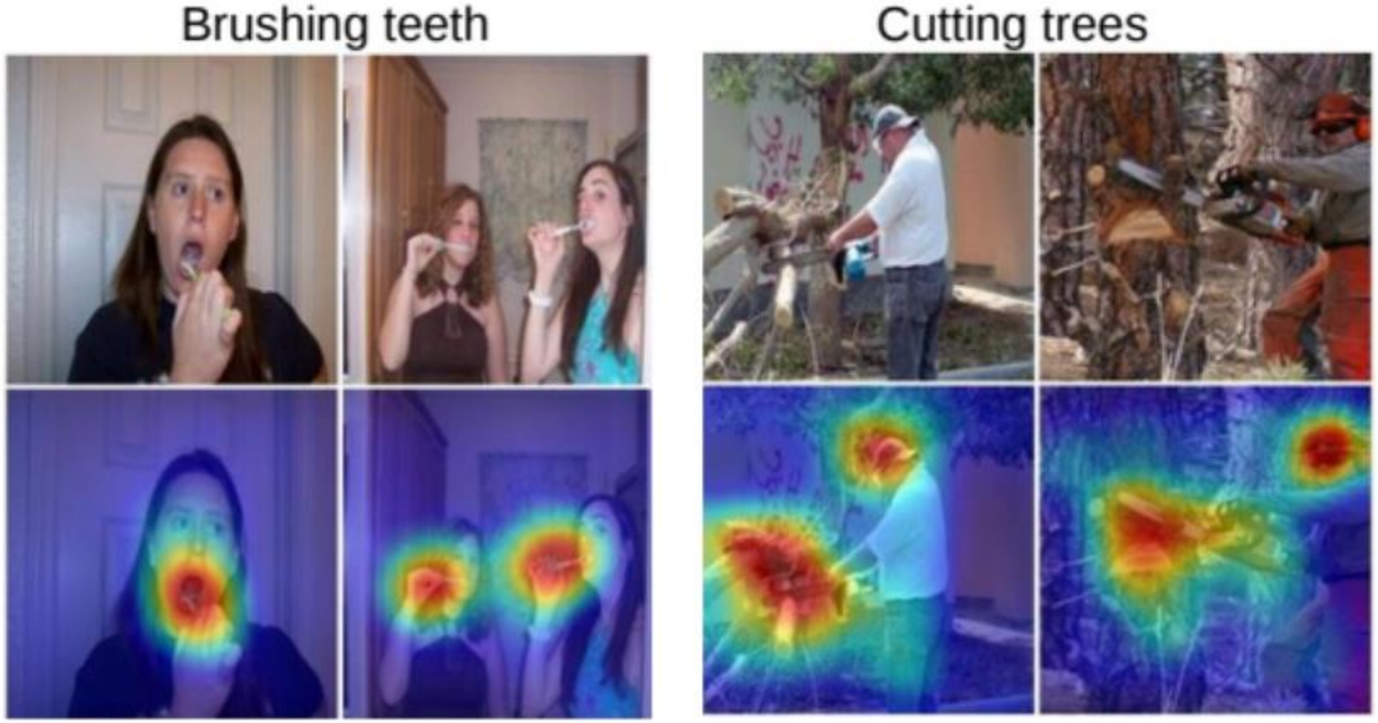
How to explain a model?





Grad-CAM

Can we find the key feature pattern like below?

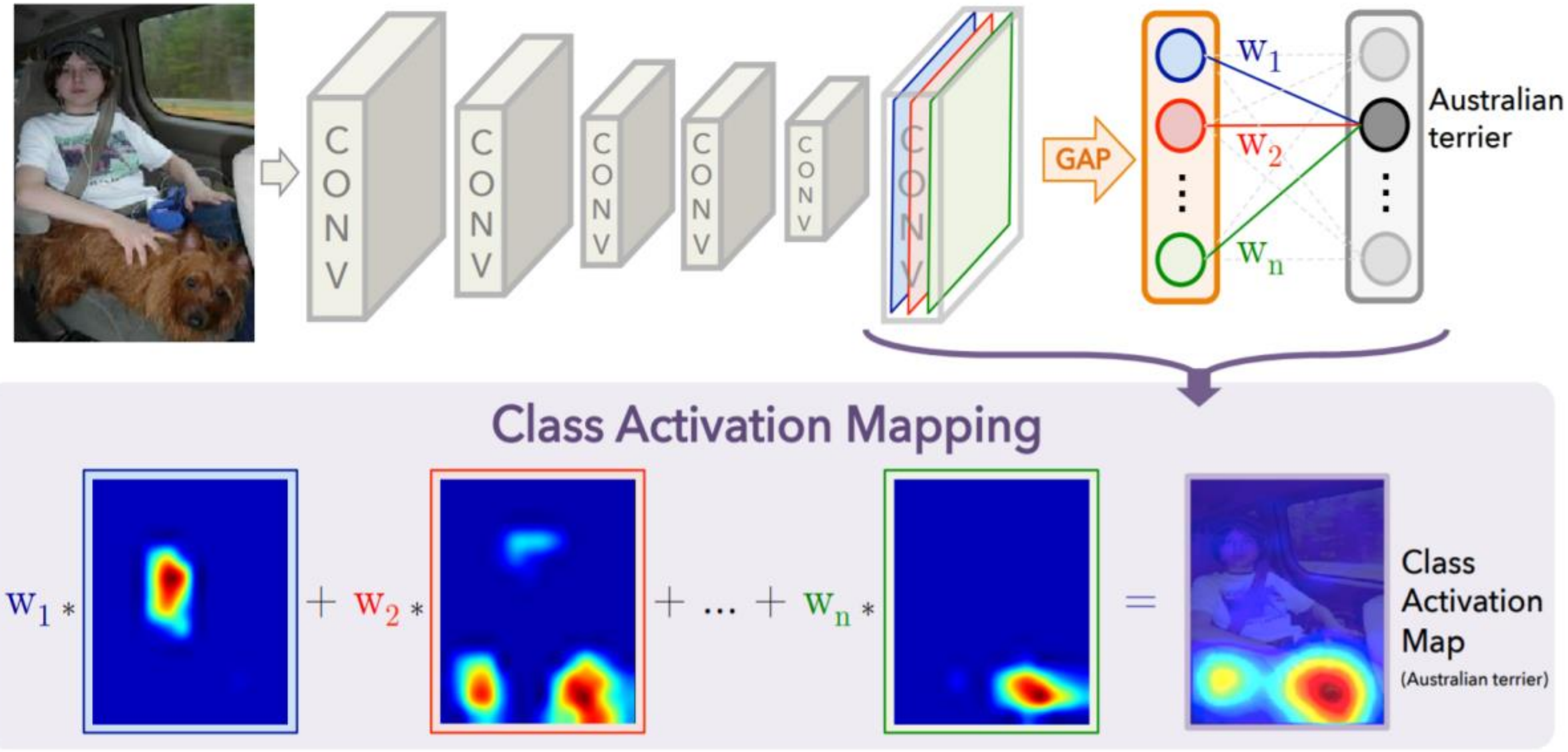


https://arxiv.org/abs/1512.04150

Grad-CAM

The GAP averages the activations of each feature map and concatenates and outputs them as a vector.

Then, a weighted sum of the resulted vector is fed to the final softmax loss layer。



How to reflect importance in GNN?

Why graph?

- Increasing the level of trust of GNN.
- Improving the transparency of the model, making the prediction more fairness, privacy and security.
- Reducing the risk of systematic errors in the model.

- Which input edges are more critical and contribute the most to the prediction?
- Which input nodes are more important?
- Which node features are more important?
- What graph patterns will maximize the prediction of a certain class?

GNNexplainer

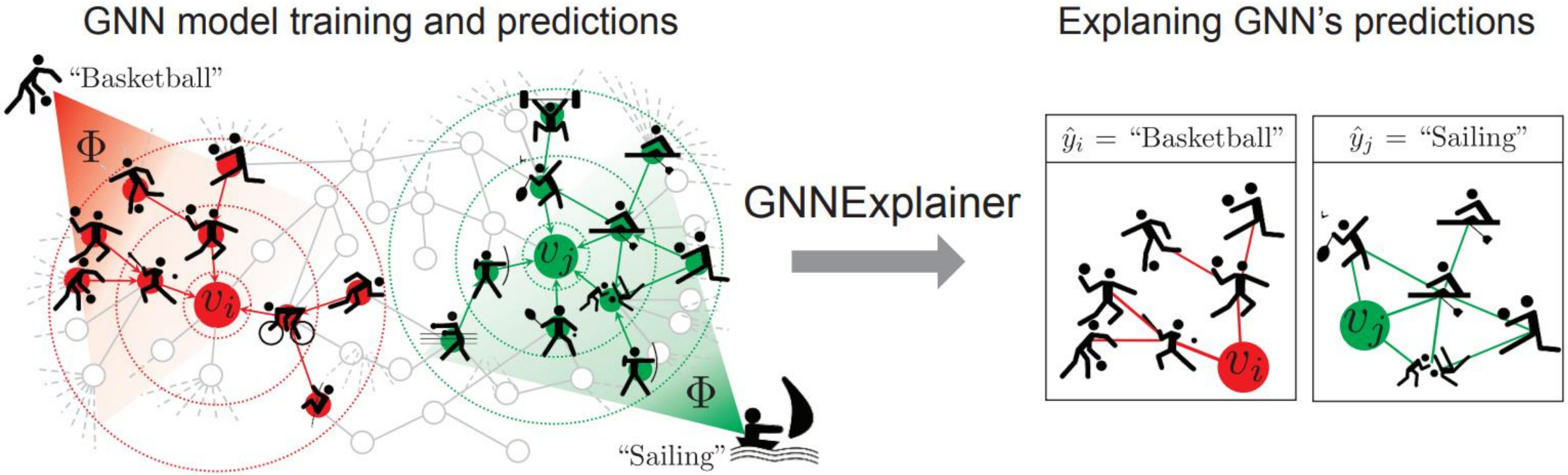
- The first general-purpose, model-agnostic interpreter for GNN models.

- The optimization task of maximizing mutual information.

- Extracts important sub-graph structures and subsets of node features for model interpretation.

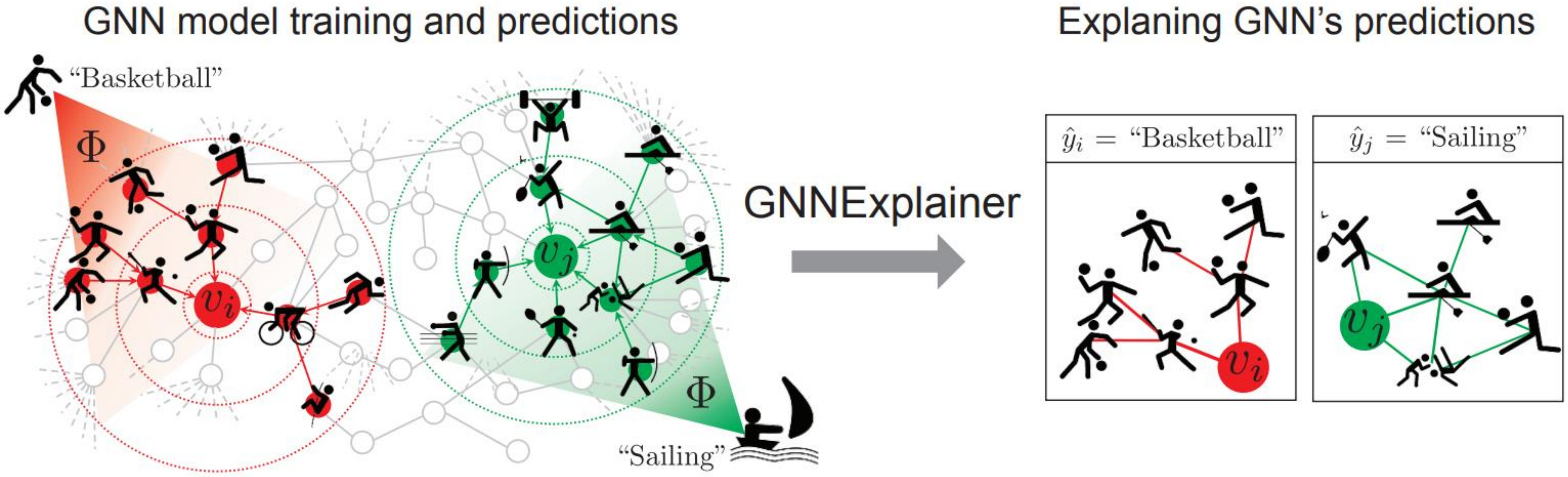
It’s a kind of application!

How to explain a GNN?



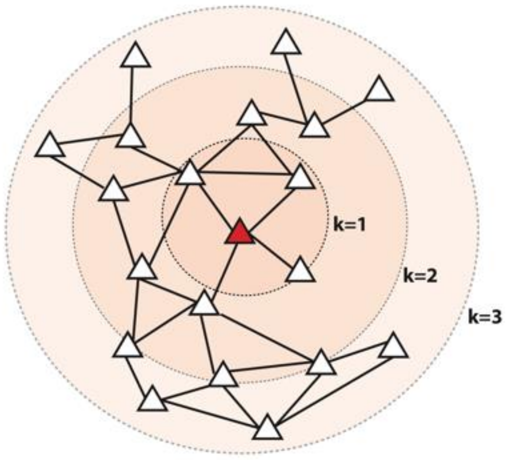
- Grad-based methods.
- Attention.

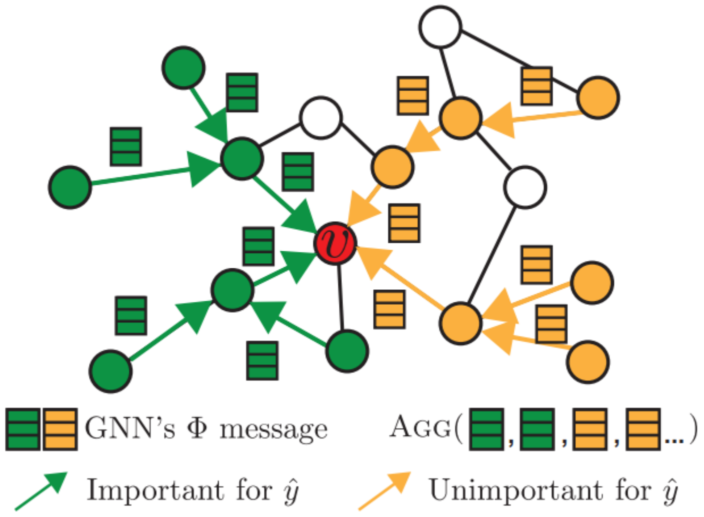
GNNexplainer



GNNExplainer的主要目标是生成一个最小图来解释一个节点或一个图的决定。这个问题可以被定义为在计算图中寻找一个子图，使用整个计算图(G)和最小图(G_s)的预测分数的差异最小。

GNNexplainer





对于一个单一节点的解释，计算图是它的k-hops邻居，其中k是模型中的卷积数。

对于一类节点的解释，本文建议选择一个参考节点并使用相同的方法来计算解释。参考节点可以通过选择其特征最接近所有其他同一类节点的平均特征的节点来选择。

对于整个图的解释，计算图成为图中所有节点的计算图的联盟。这使得计算图等同于整个输入图。

GNNexplainer math formular

G,E,V表示图，节点特征$X = x_1, \dots, x_n \in R^d$，已经训练好的GNN用$\Phi$表示，$f$表示将节点映射到$C$个不同的类别

在模型的第l层，由以下三个信息处理步骤

- 节点对(v_i, v_j)之间的信息传递，通过两个节点在上一层的编码h_i^{l-1}, h_j^{l-1}表示$m_{ij}^l = MSG(h_i^{l-1}, h_j^{l-1}, r_{ij})$。
- 对于某个节点v_i，假设其邻居为N_{v_i}，那么信息汇聚方式为$M_i^l = AGG(M_{ij}^l | v_j \in N_{v_i})$
- 进行编码更新$h_i^l = UPDATE(M_i^l, h_i^{l-1})$

GNNexplainer math formular

对于一个给定的预测结果$\hat{y}$，找到其解释(G_S, X_S^F)，其中G_S是子图，
X_S^F是子图节点对应的特征，$X_S^F = x_j^F|v_j \in G_S$。

以单个点的解释为例，转化为使子图的信息和计算图的互信息最大。

$$\max_{G_S} MI\left(Y,(G_S,X_S)\right) = H(Y) - \boxed{H(Y|G=G_S, X=X_S).}$$

等价于最小化后面一项

$$H(Y|G{=}G_S, X{=}X_S) = -\mathbb{E}_{Y|G_S,X_S}\left[\log P_\Phi(Y|G{=}G_S, X{=}X_S)\right].$$

借助于Jesen不等式（需要凸函数）

$$\min_{\mathcal{G}} H(Y|G = \mathbb{E}_{\mathcal{G}}[G_S], X = X_S).$$

(53条消息) 万能的图神经网络解释器 GNNExplainer_weixin_45519842的博客-CSDN博客
How to Explain Graph Neural Network — GNNExplainer | by Shanon Hong | Towards Data Science

GNNexplainer

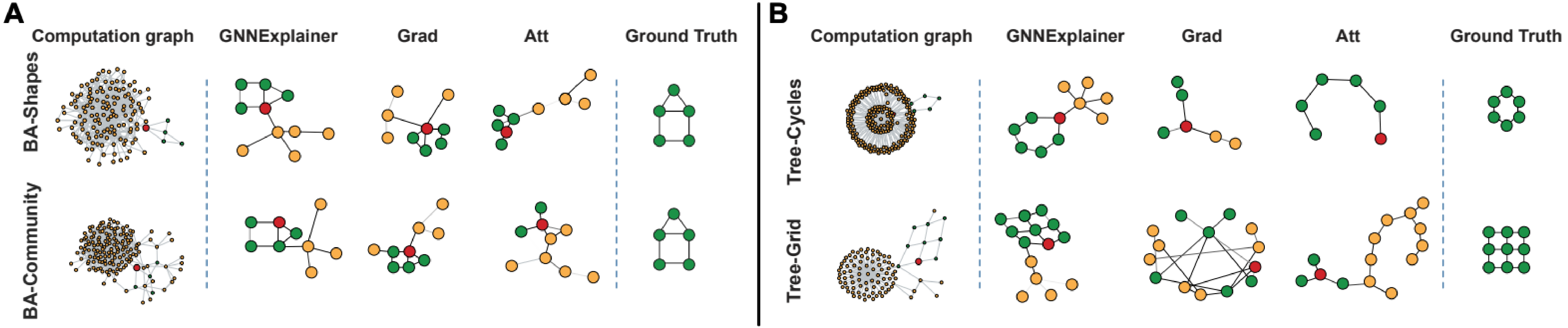


Figure 3: Evaluation of single-instance explanations. **A-B.** Shown are exemplar explanation subgraphs for node classification task on four synthetic datasets. Each method provides explanation for the red node's prediction.

GNNexplainer

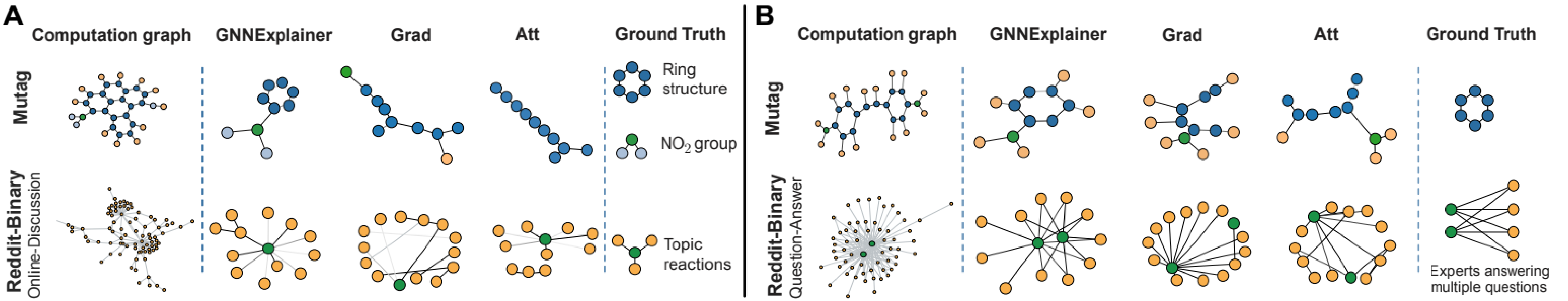


Figure 4: Evaluation of single-instance explanations. **A-B.** Shown are exemplar explanation subgraphs for graph classification task on two datasets, MUTAG and REDDIT-BINARY.